

# **SCB252 Fundamentals of Biotechniques Lab Manual**

Open Educational Resources (OER)

Na Xu and Hanqi Ye

Natural Sciences Department  
LaGuardia Community College

## Lab 1 Lab Safety, Research Integrity, Responsible Conduct of Research. (Lecture only)

Na Xu

### Objective

- To be familiar with lab safety, research integrity, and basic principles of responsible conduct of research.
- To complete Responsible Conduct of Research training through the Collaborative Institutional Training Initiative (CITI).

### Introduction

*Research integrity* includes the use of honest and verifiable methods in proposing, performing, and evaluating research, reporting research results with particular attention to adherence to rules, regulations, and guidelines, and following commonly accepted professional codes or norms. During this class, students will learn the basic principles of **the Responsible Conduct of Research** (Steneck and Zinn, 2006):

<https://ori.hhs.gov/education/products/RCRintro/index.html>

CUNY subscribes to the Collaborative Institutional Training Initiative's (CITI) research compliance courses (RCR) to fulfill CUNY's educational requirements. All CUNY faculty members, staff, postdoctoral scholars, and graduate and undergraduate students involved in research are required to complete the **CITI RCR training** within six weeks of initiating their research. RCR training certificate will be valid for four years. CUNY researchers are required to take a refresher CITI RCR training course every four years. Introduction to the RCR online course can be found here: <https://about.citiprogram.org/series/responsible-conduct-of-research-rcr/>

### Instructions to complete the RCR training through CITI:

1. Login (or register if you are a first-time user) at <https://www.citiprogram.org/Default.asp>
2. Click on "Add a course or update your learner groups"
3. Select the curriculum "Responsible Conduct of Research"
4. Select the learner group "CUNY Researchers" as your role in research.
5. Complete the required modules
6. Save a copy of your completion certificate.

### Assignment 1

Students will spend some of the class time completing the Responsible Conduct of Research (RCR) training through the Collaborative Institutional Training Initiative (CITI). A certificate will be generated upon completion of the training. Students will submit the certificate to mark the completion of this assignment.

## Lab 2 Introduction to Use *Drosophila* as a Model System for Scientific Research

Na Xu

**Objectives:** This lab provides basic information about the *Drosophila* life cycle, maintenance and handling in the laboratory, and basics in fly geneticists.

### Introduction I

The fly *Drosophila melanogaster* is one of the most intensively studied organisms in biology and serves as a model system for investigating many developmental and cellular processes common to higher eukaryotes, including humans. Of the ~1500 known *Drosophila* species, ***D. melanogaster***, the Pomace fly or fruit fly has been the most extensively used and versatile experimental model. C. W. Woodworth first cultivated the fruit fly in 1900 at Harvard Zoological laboratory. Soon, *D. melanogaster* became a popular model system for research, due to its unique features: 1. It is less expensive to rear in the laboratory in large numbers. Its small body size requires little space and facilitates handling. 2. It has a short generation time (less than 2 weeks) with large numbers of progeny. 3. It has sharply defined body phenotypes that facilitate the identification of numerous mutants. 4. Small genomic size.

*Drosophila* Life Cycle: *Drosophila* has distinct stages of development including egg, larva, pupa, and adult (Figure 1). At 25°C, an adult emerges in about 10 days. The egg of *Drosophila melanogaster* is white, about 0.5 mm in length, and laid in the food. The tiny larva hatches from the egg. The larval period consists of three motile stages or instars, the first, second, and third instars. The different larval instars can be identified by their age and body size, spiracles, and mouth hooks (Figure 1). The 1st, 2nd, and early 3rd instar larvae largely remain within the food, actively feed, and increase their body mass. Late third instar larvae come out of food and crawl in search of a dry place for pupation, which is usually the wall of a vial/bottle where they are grown in the laboratory. Larvae adhere to a dry place using the glue protein secreted by their salivary glands (Figure 1). The pupa is formed within the last larval cuticle. Gradually the cuticle hardens and becomes darker in color (Figure 2).

During the next 4-5 days, the pupa undergoes complete metamorphosis within the pupal case. Most larval tissues undergo histolysis during the metamorphosis while the adult structures differentiate (Figure 2). At the end of pupal metamorphosis, the adult fly forces its way through the anterior end of the pupal case. Freshly eclosed fly is elongated, lighter in color, and with unexpanded wings. A small dark-greenish mark of 'meconium', representing the decaying larval gut, is visible on the ventral side through the young fly's abdominal wall (Figure 1). These characteristic features of freshly emerged flies are used to isolate virgin females when needed for setting up a desired genetic cross. Within 2-3 hours, the wings expand; the body darkens. Female flies become ready to mate in 4-6 hours after eclosion.

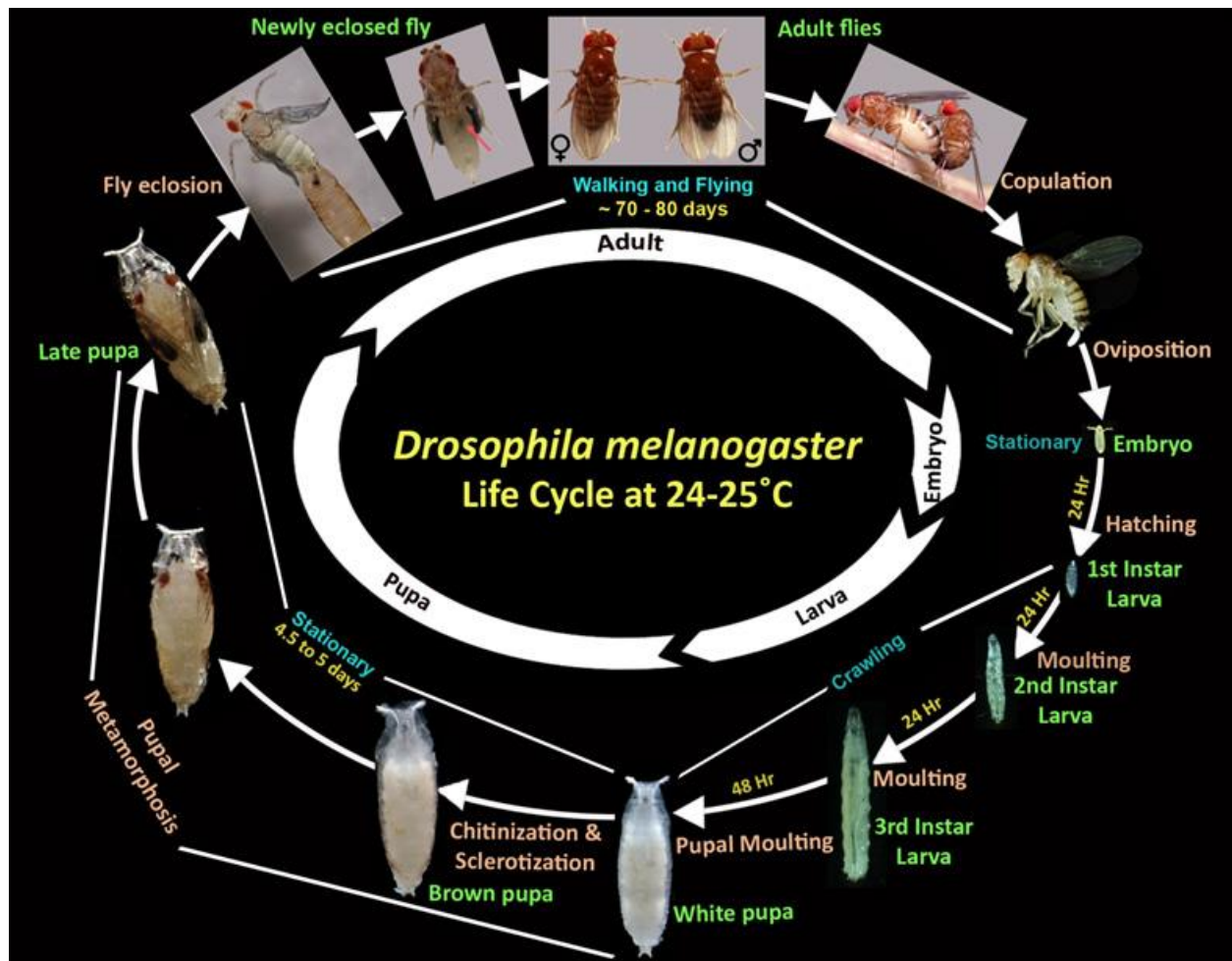


Figure 1: *Drosophila* Life Cycle. This figure shows the different stages of the *Drosophila* life cycle. The red arrow points to the 'meconium' in the newly emerged fly (S. C. Lakhota and H. A. Ranganath, 2021).

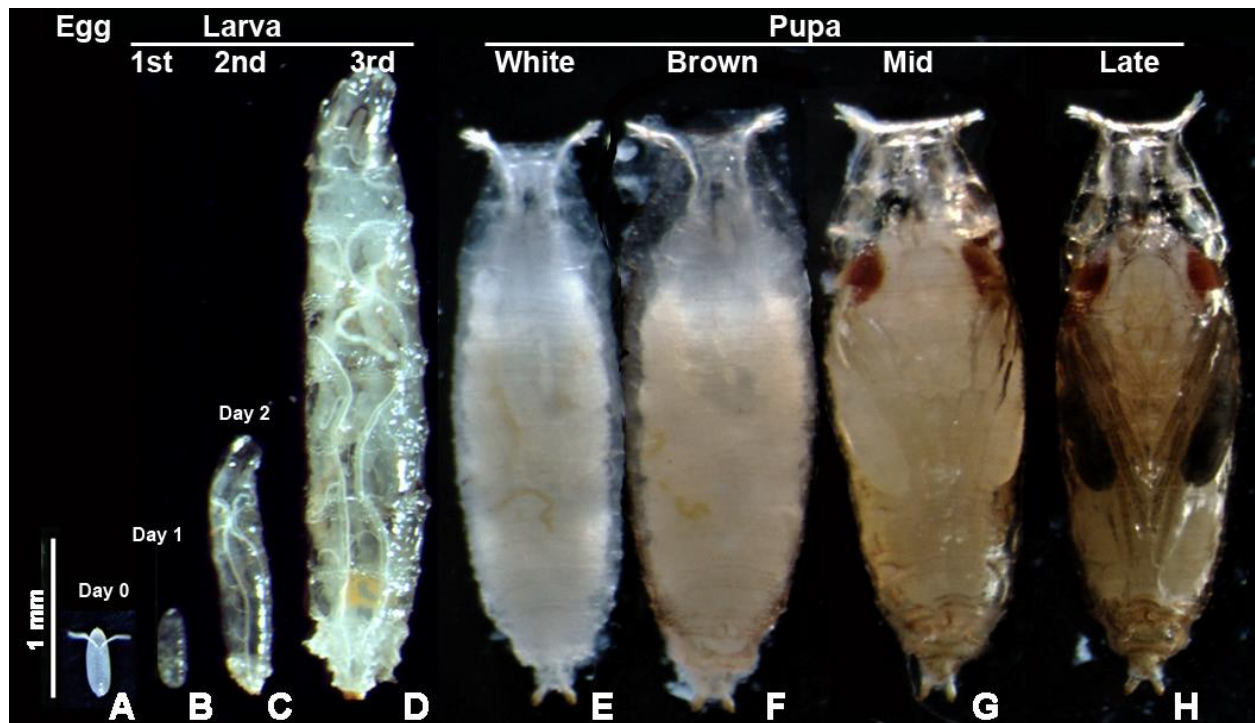


Figure 2: Larva and Pupa Stages of *Drosophila* (S. C. Lakhotia and H. A. Ranganath, 2021).

Identification of Sex: Male and female *D. melanogaster* flies (Figure 3) can be distinguished by the following features: 1. The abdomen of females has eight segments, with seven being visible. In males, the last three abdominal segments are fused, so that only five segments are visible. 2. The abdominal tip in males is considerably darker than in females of the species. 3. The male and female have different genitalia, visible on the ventral surface of the abdominal tip, with the male genital plate appearing rougher and darker. 6. The males have sex combs (Fig. 3 E') as a row of about ten stout, black bristles on the first tarsal segment of the first pair of legs.

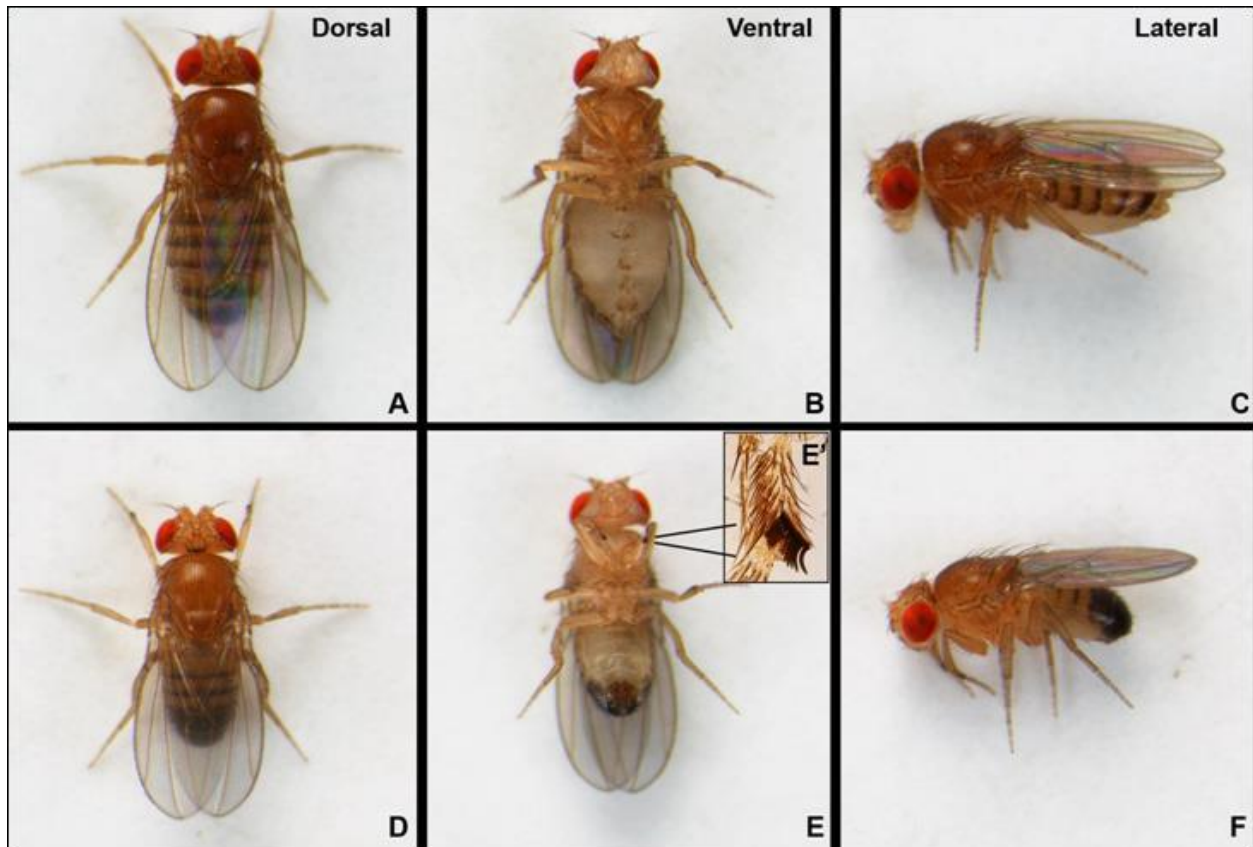


Figure 3: Female and Male Adult *D. melanogaster*. Wild type *D. melanogaster* female (A-C) and male (D-F) flies as seen from dorsal, ventral, and lateral views. The inset in E (E') shows a magnified view of the sex comb on the 1st tarsal segment of the prothoracic pair of legs (S. C. Lakhotia and H. A. Ranganath, 2021).

Stereo Microscope: The **stereo** or **dissecting microscope** is designed for low magnification observation of a sample, typically using light reflected from the surface of an object rather than transmitted through it. The instrument uses two separate optical paths with two objectives and eyepieces to provide slightly different viewing angles to the left and right eyes. This arrangement produces a three-dimensional visualization of the sample being examined. The stereo microscope is used to study the surfaces of solid specimens or to carry out close work such as dissection, microsurgery, or inspection.

### Lab Exercise I - Separating Male and Female *Drosophila*

#### Materials

Wild-type *D. melanogaster* males and females culture in a vial  
 White papers  
 Brushes  
 Fly-Nap (in the chemical hood)  
 Stereo microscope (for individual use)

### Methods

1. Students work individually on their stereo microscopes and get familiar with the operation of stereo microscopes.
2. Each student is given a few live “wild-type” cultures containing male and female flies and a piece of white paper to hold flies.
3. Students will observe the fly culture and identify different stages during fly development, such as egg, larva, pupa, and adults.
4. “Fly-Nap” is used to anesthetize flies.
5. Observe anesthetized flies under the stereo microscope, compare and contrast female and male flies, and separate female from male flies
6. Take a picture and record this as part of the lab report

### Lab Exercise Question

1. Describe the difference between female and male flies using your words.
2. In the fly culture given to you, how many are females and how many are males?

### **Introduction II**

Rearing flies in the laboratory: Ideally, flies are reared and handled in rooms that maintain a temperature between 20 - 25°C. The higher or lower temperature has adverse effects on fly viability and fecundity. Continued exposure to 30°C or above may cause sterility or death, whereas, temperatures lower than 18°C delay the life cycle and may impair viability. Appropriate food medium in suitable containers is required for maintaining *Drosophila* in the laboratory. Semi-transparent plastic bottles and vials are plugged with non-absorbent cotton or synthetic foam plugs.

Flies need care to maintain good health. Overcrowding is a common concern since flies keep laying eggs and expand their numbers within the limited space and resources available in the bottle/vial. Furthermore, the build-up of dead flies adversely affects the fitness of flies. Decaying food and fly debris attract infecting agents like bacteria, fungi, and mites. Therefore, the flies must be transferred regularly to fresh bottles/vials with food to keep them healthy. Starting a new culture with about 10-20 flies in vials and <50 flies in bottles avoids overcrowding and maintains a healthy culture. As a good practice, flies reared at 25°C must be transferred to fresh food within 25-30 days if cultured in bottles and 15-20 days if cultured in vials.

Care should be taken to avoid flies escaping from the containers during the maintenance process. These flies may contaminate other growing cultures in the laboratory or fresh food vials/bottles. To get rid of such unwanted flies in the lab, Fly traps should be set up in the lab area. A simple Fly trap can be prepared with a fresh food bottle with a funnel placed at its open



end. A dilute solution of activated yeast with a few drops of propionic acid as a thin fluid layer on the food surface attracts the unwanted escapee flies to enter the trap through the funnel. Replace the fly traps weekly, if not earlier.

*Killing Flies and Setting up a Morgue:* This is an unfortunate necessity when using flies. A bottle or beaker with soapy water or mineral oil is generally used. Dump anesthetized flies directly into the soapy water or mineral oil where they drown. A bottle (beaker, or screw-capped jar) filled with ethanol or isopropanol can also be used as a morgue.

## **Lab Exercise II - Preparing Fly Food and Fly Trap**

### Materials

Hotplate  
Nutri-Fly food  
Distilled water  
Balancer  
1 L and 100 mL beakers  
Spatula  
Microwave  
Thermometer  
Heat resistant gloves  
Tegosept  
Ethanol  
Phosphoric Acid  
Fly food bottles, vials, and trays  
Cotton and foam plugs  
Funnel  
Live yeast dilution

### Methods

1. Preheat the hotplate at setting 3
2. Weigh 89 grams of Nutri-Fly food
3. While the plate is heating, place a 1 L beaker on the plate, add 500 mL of distilled water, and slowly pour in the 89 g of Nutri-Fly food. Use a spatula constantly to break up any food bubbles or lumps.
4. Microwave for 2 min and then stir with a spatula. Repeat the second time for a total of 4 min.
5. Microwave for an additional 30-45 seconds without letting the food boil over.

*(Watch the meniscus carefully. When it starts to rise, stop the microwave right away)*

6. Place the beaker on a hotplate, cover with foil, and stir. The food should be at the boiling stage, making some bubbles steadily, but not boiling over the container. During this time, insert a thermometer through the foil cover such that the bulb is just below the surface but clear of the stir bar. Set up a timer for 10 min.
7. During the 10 min waiting, prepare a 10% Tegosept (dissolve 0.5 g Tegosept in 5 mL ethanol). Tegosept is an anti-fungal drug to prevent fungi infection in fly cultures.
8. After 10 min, put the food on the bench. Keep stirring until the temperature drops to 75 °C.
9. Remove the thermometer, add 5 mL of 10% Tegosept and 170 uL of Phosphoric Acid and mix thoroughly by hand using a spatula. The fly food is ready and needs to be poured into containers quickly before they solidify. Phosphoric Acid is used to acidify food and has fungicidal effects. Phosphoric Acid is a strong acid and corrosive. Gloves are needed when handling Phosphoric Acid.
10. Put small aliquots into a 100 mL beaker and pour from the smaller beaker. Students work as 3-4 groups and pour 100 mL of food into containers. Pour about 5 ml per vial or 30 ml per bottle. Wipe excess food off the sides of the vial/bottle with tissue paper.
11. Stack the trays onto each other or cover the top tray with paper towels to keep flies from the food before you can plug.
12. Leave at room temperature overnight to dry.
13. The next day, plug in cotton and place food in plastic bags. Store at room temperature for short-term use and in the refrigerator for long-term use (the lab technician or instructor will complete this)
14. Use a fresh food bottle with a funnel placed at its open end as a fly trap. Put a dilute solution of activated yeast with a few drops of phosphoric acid as a thin fluid layer on the food surface to attract unwanted flies in the room (this will be completed in the following lab).

## **Assignment 2:**

Students will read the beginning chapters of [\*"Genetics on the Fly: A Primer on the Drosophila Model System"\*](#) (Hales et al., 2015) including the chapters in "Natural History", "Life Cycle", "Food and Husbandry" and "Sex Determination", and summarize the reading in their own words to

introduce *Drosophila* as a model system to conduct research. The minimum word requirement for this assignment is 500 words.

### Lab 3 Classic *Drosophila* Mutations and Their Phenotypes

Na Xu

**Objective:** This lab provides basic information about the *Drosophila* adult anatomy, genes, mutations, and phenotypes, and basics in fly geneticists. Students will familiarize themselves with common visible mutations affecting different parts of *D. melanogaster* and with the genetic basis of these mutant phenotypes.

#### Introduction

Wild Type *Drosophila*: The 'wild type' is ascribed to the most common phenotype seen in natural populations. Various natural fly populations have been collected from different geographical areas by researchers at different times and maintained in the laboratory as distinct 'wild type' populations since the original time of collection.

*Drosophila* chromosomes: *Drosophila melanogaster* has four pairs of chromosomes, which include a pair of sex chromosomes (a pair of X chromosomes in females and the X and Y pair in males) and three pairs of autosomes (chromosomes 2, 3, and 4) (Figure 1). The 4th chromosome is very small and often named as 'dot' chromosome (Figure 1). The two arms of the long X, 2nd, and 3rd, chromosomes are designated as left (XL, 2L, 3L) and right arms (XR, 2R, 3R) while those of the Y chromosome are designated as long (YL) and short (YS) arms (Figure 1).

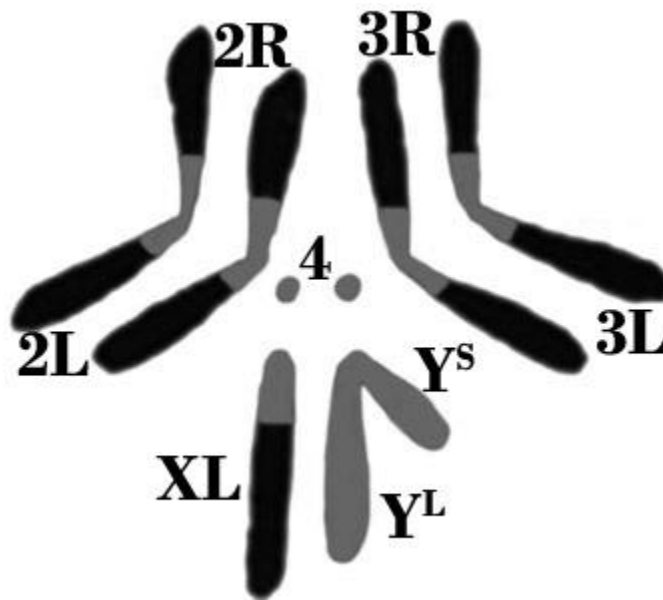


Figure 1: Schematic of Mitotic Chromosomes of *Drosophila Melanogaster*. L and R following the chromosome name (X, 2, and 3) indicate left and right arms, respectively; the very short right arm of the acrocentric X is not distinguishable. YS and YL represent the short and long arms of the Y-chromosome, respectively. The 'dot' chromosome 4 does not show left or right arms.

Light and dark shaded regions represent heterochromatic and euchromatic regions, respectively (S. C. Lakhota and H. A. Ranganath, 2021).

*Anatomy of Female and male adult flies:* Females (Figure 2A) have a pointed abdomen whereas males (Figure 2B) have a rounded abdomen. Males have greater and more uniform pigmentation of the abdominal tip. Females are typically larger than males, with light-colored, pointed genitalia, which is more rounded and rougher than males. Males characteristically exhibit a small patch of dense bristles on the forelegs, called the sex comb, which aids in grasping the female during mating.

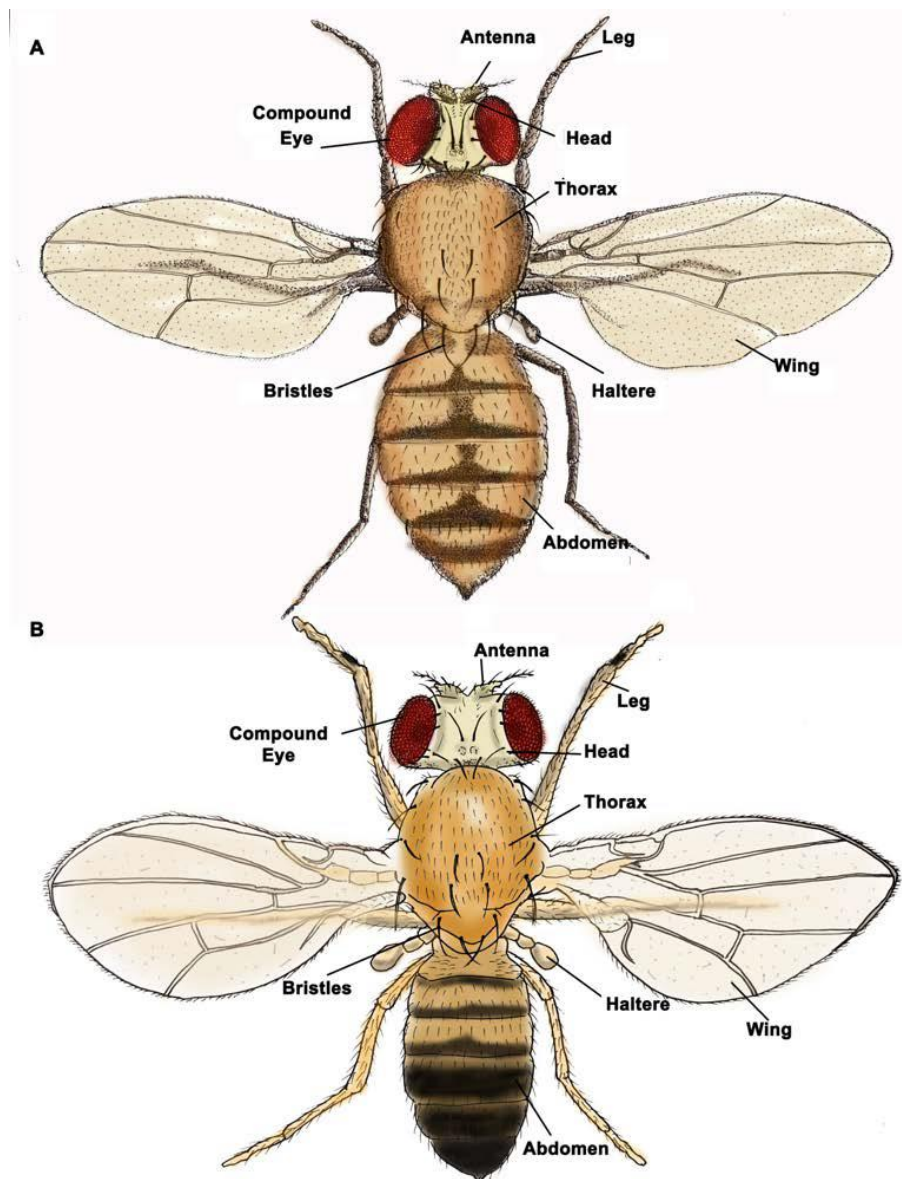


Figure 2: Anatomy of Wild Type Adult Female (A) and Male (B) *Drosophila Melanogaster* (S. C. Lakhota and H. A. Ranganath, 2021).

Classic mutants and their phenotypes:

**Bar eye mutation:** Gene name: Bar; Symbol: B; Affected chromosome: X; Semi-Dominant. The ommatidial array in eyes is restricted to a narrow vertical bar-like area with only about 90 facets in the male and 70 facets in homozygous females, in contrast with the normal numbers of about 740 facets in males and 780 in females. Heterozygous females show a 'kidney-shaped' eye with about 360 facets in each eye (Figure 3). This mutation is due to a tandem duplication of a small part (16A1-7 chromosome region) of the X-chromosome.

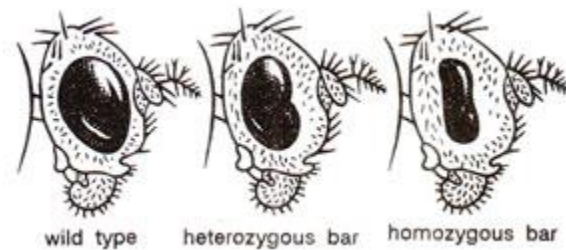


Figure 3: Shapes of eyes in wild-type, heterozygous bar, and homozygous bar flies ([https://biocyclopedia.com/index/genetics/structural\\_changes\\_in\\_chromosomes/duplications.php](https://biocyclopedia.com/index/genetics/structural_changes_in_chromosomes/duplications.php))

**Eyeless mutation:** Gene name: eyeless; Symbol: ey; Affected chromosome: 4; Dominant mutation. These flies have no eyes (Figure 4). They have a defect in their "eyes absent" gene, which normally instructs cells in the larvae to form an eye.



Figure 4: Drosophila Eyeless Mutation. (<https://www.carolina.com/flies/drosophila-living-eyeless-chromosome-4-mutant/172620.pr>)

**Vestigial wings mutation:** Gene name: vestigial; Symbol: vg; Affected chromosome: 2; Recessive mutation. In the classical vg mutants, the wings of homozygotes are reduced to vestiges and are usually held at right angles to the body; wing veins still visible (Figure 5C).

Temperatures of 29°C or greater appreciably increase wing size. Flies with vestigial wings cannot fly.

**Ebony body mutation:** Gene name: ebony; Symbol: e; Affected chromosome: 3; Recessive mutation. Body color varies from shining black to slightly darker than wild-type flies (Figure 5B). Normally, the ebony gene is responsible for building up the tan-colored pigments in the normal fruit fly. If the ebony gene is defective, the black pigments accumulate all over the body.

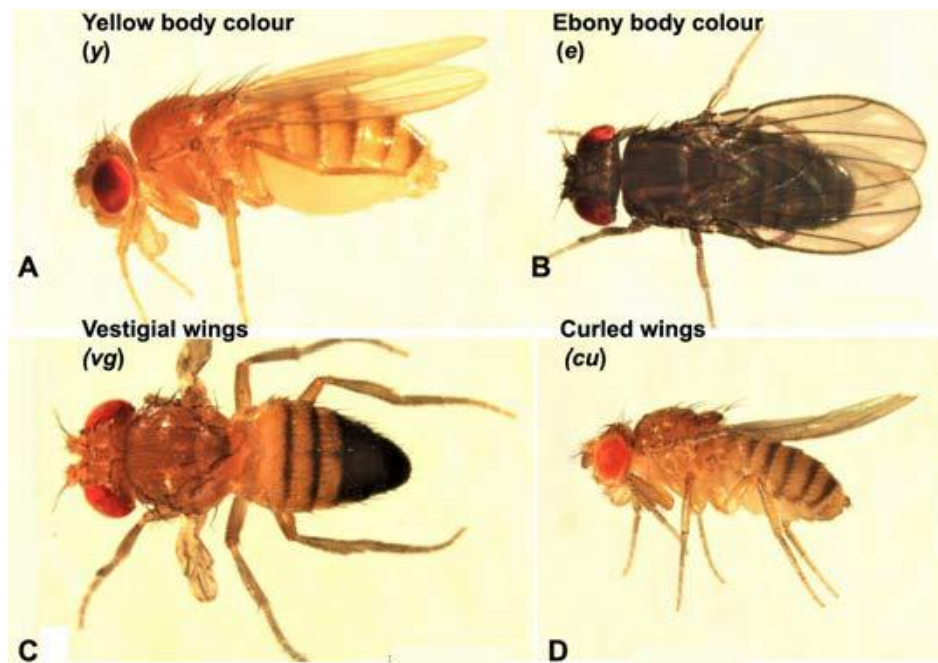


Figure 5: Mutations Affecting Body Colour or Wing Size/Shape (S. C. Lakhotia and H. A. Ranganath, 2021).

Collection of virgins: Many experiments require crosses between males and females of specifically desired genotypes. *Drosophila* female stores sperms transferred during the mating in their spermatheca and use them lifelong. Therefore, it becomes imperative to select unmated virgin females to set crosses between flies of the desired genotypes. Female flies mate only 4-6 hours after eclosion at 25°C. Therefore, segregation of males and females in separate vials/bottles within a window of 4 hours after emergence ensures that the females remain virgins.

### **Lab Exercise - Observe Mutations and Prepare for Genetic Crosses to generate flies carrying double mutations**

#### Materials

Fly cultures with wild-type flies and mutant fly with a mutation in *vestigial*, *eyeless*, *bar* or *ebony*  
Stereo microscopes  
Fly food in bottles and viable  
Markers

Fly-nap  
Brushes and forceps  
White papers

### Methods

1. Students are given wild-type flies (Figure 6) and different mutant flies to examine and identify the mutant characters externally visible and to match them with their phenotypes.
2. Students record their observations in writing and pictures.



Figure 6: Wild Type Female (A) and Male (B) Flies as Seen in Lateral View (S. C. Lakhotia and H. A. Ranganath, 2021).

3. Students design their genetic crosses to generate a fly carrying double mutations
4. Students will collect 5 virgins from one mutant and mate them with 5 males from another mutant. Transfer the cross into a new vial, and label the cross and the date.
5. Leave the cross at room temperature for two weeks in order to observe F1 progeny and select flies carrying double mutations.

### **Assignment 3:**

Students will describe which flies were observed and their features compared to the wild-type flies. Students should include a picture of the wild-type and mutant flies in their assignment. Students will write their genetic crosses and plans to generate a double mutant fly.



## Lab 4 Isolating Genomic DNA from Fly

Na Xu

**Objective:** The lab provides basic information about genomic DNA and technology to isolate *Drosophila* genomic DNA

### Introduction

**Genomic DNA:** A cell's DNA, packaged as a double-stranded DNA molecule, is called its genome. Genomic DNA **constitutes the total genetic information of an organism**. Genomic DNA molecules are generally large, and in most organisms are organized into DNA-protein complexes called chromosomes. In research, genomic DNA is a useful tool in applications such as PCR, library construction, Southern blotting, hybridizations, SNP analysis, and molecular diagnostic assays.

**Genome Sequence of *Drosophila*:** The genome of *Drosophila melanogaster* is about 180 megabase (Mb) in size, with roughly two-thirds euchromatic and one-third heterochromatic (Figure 1). The euchromatin contains about 98% of the protein-coding genes in the genome; the heterochromatin is largely composed of simple sequence repeats (Figure 1) (Celniker and Rubin, 2003).

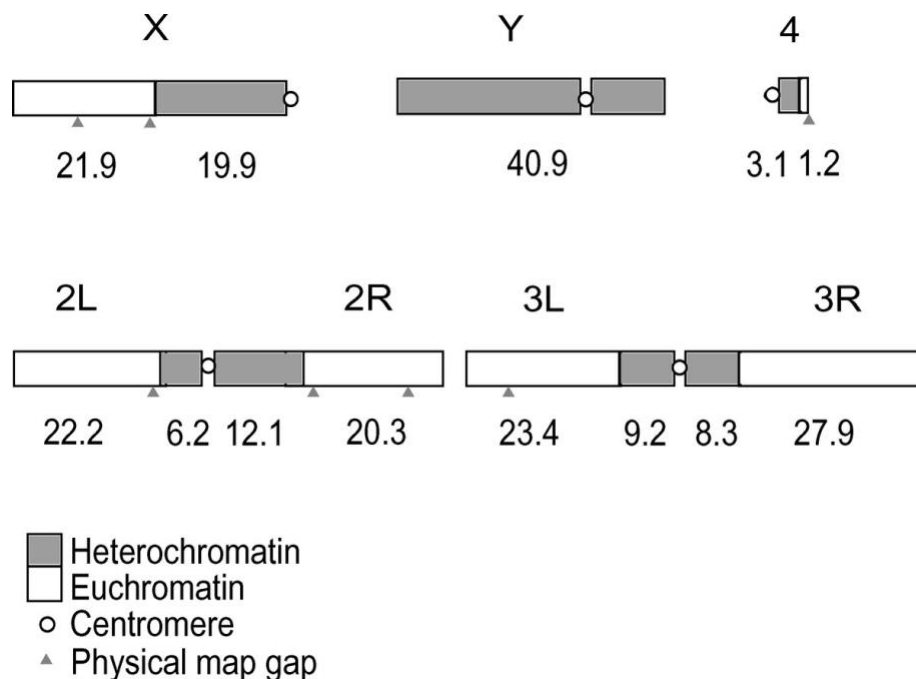


Figure 1: The Chromosomes of *Drosophila melanogaster* (Celniker and Rubin, 2003).

**Quantification of DNA:** Reliable measurement of DNA concentration is important for many applications in molecular biology. Several methods are commonly used to measure both genomic and plasmid DNA concentrations.

1. **Spectrophotometry** can be used to measure microgram quantities of pure DNA samples (i.e., DNA that is not contaminated by proteins, phenol, agarose, or RNA). DNA concentration can be determined by measuring the absorbance at 260 nm ( $A_{260}$ ) in a spectrophotometer using a quartz cuvette. For the greatest accuracy, readings should be between 0.1 and 1.0. An absorbance of 1 unit at 260 nm corresponds to 50  $\mu$ g genomic DNA per ml ( $A_{260}=1$  for 50  $\mu$ g/ml; based on a standard 1 cm path length. This relation is valid only for measurements made at neutral pH, therefore, samples should be diluted in a low-salt buffer with neutral pH (e.g., Tris·Cl, pH 7.0). Spectrophotometric measurements do not differentiate between DNA and RNA, so RNA contamination can lead to an overestimation of DNA concentration. The **NanoDrop Spectrophotometer** is a spectrophotometer that enables highly accurate analyses of 1  $\mu$ l samples for DNA, RNA, protein, pigments, whole UV-visible (220-700nm) spectrum, and cell density tests with remarkable reproducibility.
2. **Fluorometry** is more sensitive, allowing the measurement of nanograms of DNA. Hoechst 33258 dye has little affinity for RNA, allowing accurate quantification of DNA samples that are contaminated with RNA. It shows increased emission at 458 nm when bound to DNA. DNA standards and samples are mixed with Hoechst 33258 and measured in glass or acrylic cuvettes using a scanning fluorescence spectrophotometer or a dedicated filter fluorometer set at an excitation wavelength of 365 nm and an emission wavelength of 460 nm. The sample measurements are then compared to the standards to determine DNA concentration.
3. **Agarose gel analysis** enables quick and easy quantification of DNA, especially for small DNA fragments (such as PCR products). As little as 20 ng DNA can be detected by agarose gel electrophoresis with ethidium bromide staining. The DNA sample is run on an agarose gel alongside known amounts of DNA of a similar size. The amount of sample DNA loaded can be estimated by comparison of the band intensity with the standards using a scanner or imaging system. Be sure to use standards of roughly the same size as the fragment of interest to ensure reliable estimation of the DNA quantity, since large fragments intercalate more dye than small fragments and give a greater band intensity.

## Lab Exercise - Isolate Genomic DNA from *Drosophila*

### Materials

Quick-DNA™ Miniprep Kit (Zymo Research)  
Tissue homogenizer, micropestle, and 1.5 mL microtubes  
Wild-type fly culture  
Fly culture with ebony mutants  
NanoDrop Spectrophotometer

### Methods

1. Place 5 adult flies in a 1.5 mL microtube fitted with a micro pestle and thoroughly grind in 500 µl of **Genomic Lysis Buffer**. Mix gently by inverting the microtube. Incubate for 1 hr at 37°C.
2. Centrifuge the lysate at top speed (10,000 x g) for 5 minutes. Making sure not to disturb the pelleted debris, transfer the **supernatant** to a **Zymo-Spin™ IICR Column** in a Collection Tube and centrifuge at 10,000 x g for one minute. Discard the Collection Tube with the flow through.
3. Transfer the **Zymo-Spin™ IICR Column** to a new Collection Tube. Add 200 µl of **DNA Pre-Wash Buffer** to the spin column. Centrifuge at 10,000 x g for one minute.
4. Add 500 µl of **g-DNA Wash Buffer** to the spin column. Centrifuge at 10,000 x g for one minute.
5. Transfer the spin column to a clean microcentrifuge tube. Add 50 µl **DNA Elution Buffer** or water<sup>1</sup> to the spin column. Incubate for 2-5 minutes at room temperature and then centrifuge at top speed for 30 seconds to elute the DNA. The eluted DNA can be used immediately for molecular-based applications or stored at ≤ -20 °C for future use.
6. Quantify the DNA using **NanoDrop Spectrophotometer**, and record the quantification on DNA storage tubes. Use a single drop on the NaoDrop machine and blank the machine using distilled water.

#### Assignment 4

Students will continue reading the “**Drosophila Genome Features**” chapter of [\*Genetics on the Fly: A Primer on the Drosophila Model System\*](#) (Hales et al., 2015), and summarize the reading in their own words. The minimum word requirement for this assignment is 500 words.

## Lab 5 Genetic Crosses and Introduction to Bioinformatics

Hanqi Ye and Na Xu

### Objective:

- To be familiar with genetic crosses and the laws of inheritance
- To be familiar with the NCBI and Flybase database
- To practice and use the fundamental bioinformatics tool: BLAST

### Introduction

Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data. The main driving force for the emergence and rapid development of bioinformatics comes from the application of high-throughput technologies such as next-generation sequencing in the field of life science. At present, the genomes of humans, animals, insects, and plants have been sequenced. The number of known DNA sequences exceeds 20 trillion bases, doubling every five months. The unprecedented big data contains a lot of unknown new discoveries and new knowledge, bringing new historical opportunities to life science research.

At the largest scale, bioinformatics is divided into databases and software tools. The database is classified as large centralized resources like the **National Center for Biotechnology Information/NCBI** (<https://www.ncbi.nlm.nih.gov/>) of the United States, **European Bioinformatics Institute/EBI** (<https://www.ebi.ac.uk/>) of Europe, **Genome Browser of UCSC** (<https://genome.ucsc.edu/>), and specific independent resources built by research groups like **Flybase** (<http://flybase.org>) which is a joint project with the Berkeley Drosophila Genome Project.

**FlyBase** is the leading database and web portal for genetic and genomic information on the fruit fly *Drosophila melanogaster* and related fly species. All databases offer powerful software to analyze and interpret big biological data. One of the most useful tools is the **Basic Local Alignment Search Tool (BLAST)**. BLAST finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. This tool makes impossible missions like finding a drop in the ocean possible.

### Lab Exercise I - Observe F1 Generation Phenotypes from the Lab 3 Crosses

#### Material

Genetic crosses done in lab 3

#### Methods

1. Students revisit their crosses set up in lab 3, and observe F1 progeny and their phenotypes.

- Students count the total number of F1 progeny, separate flies based on their phenotypes, and count the number of flies in each phenotype.
- Calculate F1 phenotype and genotype ratios and describe how these observations may fit students' original plan of genetic crosses.
- Reflect on how this observation fits the law of inheritance, such as Mendel's law of segregation and independent assortment.

## Lab Exercise II - Get to know the NCBI: blast unknown genes and protein sequences

### Material

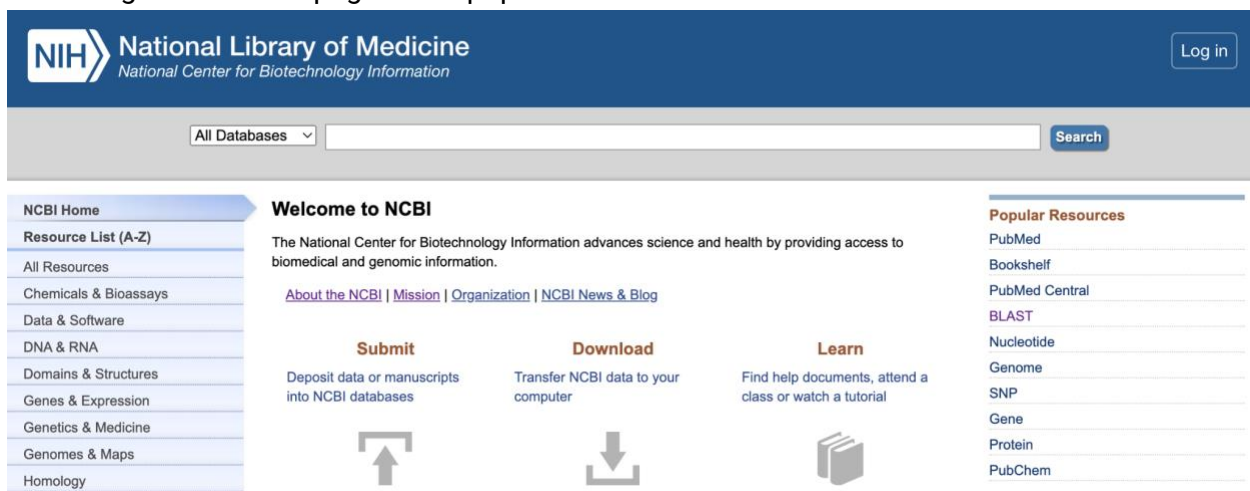
Laptops

Two unknown sequences

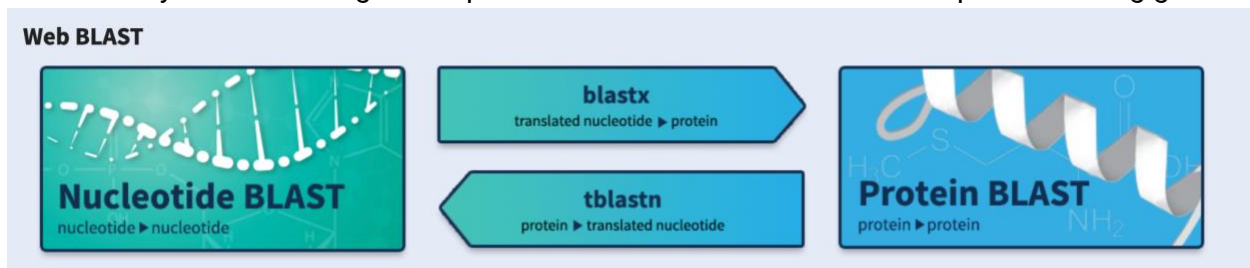
NCBI home page (<https://www.ncbi.nlm.nih.gov/>)

### Methods

- Students download two unknown sequences [here](#).
- Students open NCBI home page (<https://www.ncbi.nlm.nih.gov/>), click on BLAST at the right side of the page under popular resources.



- Use Blast to search nucleotide sequences and Blastp to search protein sequences. Blastx searches protein databases using a translated nucleotide query. Students can verify the unknown gene sequence with blastx results since it is a protein-coding gene.



- In the blast page, students upload the unknown sequences, use all default parameters, check the box of ' show result in a new window, and click on the BLAST button. The result page will refresh itself until you see the blast summary.

- Find the best hit of the query sequences. As shown below, the Blast query alignments are displayed in rank order according to their "Max Score". For this example, the top hit from the Demo scaffold sequence was Escherichia coli strain 2011C-3911 with Accession #CP015240.1.

Descriptions	Graphic Summary	Alignments	Taxonomy				
Sequences producing significant alignments							
Download Manage Columns Show 100 ?							
<input checked="" type="checkbox"/> select all 100 sequences selected							
<a href="#">GenBank</a> <a href="#">Graphics</a> <a href="#">Distance tree of results</a>							
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">Escherichia coli strain 2011C-3911, complete genome</a>	55042	56150	100%	0.0	99.85%	<a href="#">CP015240.1</a>
<input checked="" type="checkbox"/>	<a href="#">Escherichia coli strain ESKL 15 chromosome, complete genome</a>	54765	55410	100%	0.0	99.68%	<a href="#">CP041678.1</a>
<input checked="" type="checkbox"/>	<a href="#">Escherichia coli strain AMSHJX01 chromosome, complete genome</a>	54748	55393	100%	0.0	99.67%	<a href="#">CP030939.1</a>
<input checked="" type="checkbox"/>	<a href="#">Escherichia coli strain S21, complete genome</a>	54741	55856	100%	0.0	99.67%	<a href="#">CP010230.1</a>
<input checked="" type="checkbox"/>	<a href="#">Escherichia coli strain H14, complete genome</a>	54741	55856	100%	0.0	99.67%	<a href="#">CP010177.1</a>
<input checked="" type="checkbox"/>	<a href="#">Escherichia coli strain 2015C-3905 chromosome, complete genome</a>	54732	55376	100%	0.0	99.66%	<a href="#">CP027371.1</a>
<input checked="" type="checkbox"/>	<a href="#">Escherichia coli strain 2014C-3097 chromosome, complete genome</a>	54726	55371	100%	0.0	99.66%	<a href="#">CP027449.1</a>
<input checked="" type="checkbox"/>	<a href="#">Escherichia coli strain ATCC 43896 chromosome, complete genome</a>	54726	55847	100%	0.0	99.66%	<a href="#">CP024278.1</a>

- Clicking the Description name for one of the Blast hits will display a sequence alignment of the query sequence vs. the NCBI nucleotide database sequences, as shown in the below figure. This result includes the Karlin-Altschul Expect value, the percent identity of the alignment, the number of gaps in the alignment, and other pertinent information.

## Escherichia coli strain 2011C-3911, complete genome

Sequence ID: [CP015240.1](#) Length: 4863599 Number of Matches: 4

Range 1: 4671753 to 4701689 [GenBank](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
55042 bits(29806)	0.0	29896/29940(99%)	3/29940(0%)	Plus/Plus
Query 1		CCTCCCAAACCTGGCGATATCCAGTCTGGTAAACAGCCTCAAAGGCGTATCGGGTAGGTTA	60	
Sbjct 4671753		CCTCCCAAACCTGGCGATATCCAGTCTGGTAAACAGCCTCAAAGGCGTATCGGGTAGGTTA	4671812	
Query 61		CTGCGACGAGATCGACCAGATATTGCAGTCAGGTATTACTACAAAGGCGTTTTGTGGAGT	120	
Sbjct 4671813		CTGCGACGAGATCGACCAGATATTGCAGTCAGGTATTACTACAAAGGCGTTTTGTGGAGT	4671872	
Query 121		CCTGGCTATTTTGCCAGTAGCTGCGGAGGTGCGCCAATATCCGTCATCCGCCAATACATT	180	
Sbjct 4671873		CCTGGCTATTTTGCCAGTAGCTGCGGAGGTGCGCCAATATCCGTCATCCGCCAATACATT	4671932	

## Lab Exercise III - Search and download required *Drosophila melanogaster* gene sequences

### Material

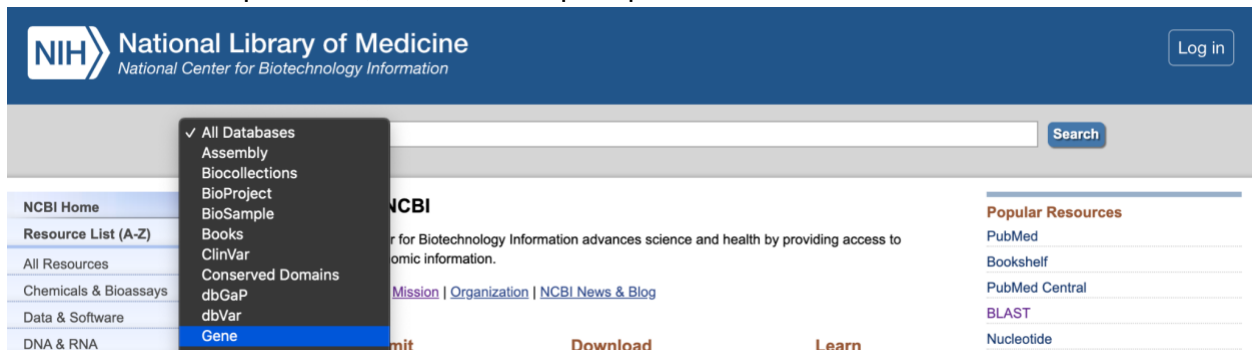
Laptop

NCBI home page (<https://www.ncbi.nlm.nih.gov/>)

Flybase (<http://flybase.org>)

### Methods

1. Students open **NCBI**, click on the top dropdown box and choose 'Gene'.



2. At the search box, search four *D. mel* genes: **Bar**, **Eyeless**, **Vestigial**, and **Ebony** respectively.



Gene   [Create RSS](#) [Save search](#) [Advanced](#)

Tabular ▾ 20 per page ▾ Sort by Relevance ▾ Send to: ▾

See [ADRB2 \(BAR\) adrenoceptor beta 2](#) in the Gene database  
bar in [Homo sapiens \(3\)](#) [All 16 Gene records](#)

**Search results**

Items: 1 to 20 of 73489 << First < Prev Page 1 of 3675 Next > Last >>

See also 1262 discontinued or replaced items.

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> <a href="#">hh</a> ID: 42737	hedgehog [ <i>Drosophila melanogaster</i> (fruit fly)]	Chromosome 3R, NT_033777.3 (23128169..23141906, complement)	Dmel_CG4637, CG4637, DmelCG4637, Dmhh, HH, Hg, Hh, Mir, Mrt, anon-WO0134654.19, anon-WO0182946.19, bar, bar-3, bar3, hg, l(3)hh, l(3)neo56, l(3)neo57	
<input type="checkbox"/> <a href="#">B</a> ID: 44798	Bar [ <i>Drosophila melanogaster</i> (fruit fly)]		Bar, BarH1, bar	
<input type="checkbox"/> <a href="#">B-H1</a> ID: 32724	BarH1 [ <i>Drosophila melanogaster</i> (fruit fly)]	Chromosome X, NC_004354.4	Dmel_CG5529, BH1, Bar, Bar H1, Bar-H1, BarH, BarHI, Barh1, CG5529,	

clear

**Filters:** [Manage Filter](#)

**Results by taxon**

Top Organisms [\[Tr\]](#)

Apaloderma vittatu  
Caenorhabditis ele  
Salmo trutta (206)  
Oncorhynchus myk  
Cyprinus carpio (2  
All other taxa (580  
More...

**Find related data**

Database:

**Search details**

Barfall Fields

- Click on the best hit and find the download datasets button. Download gene, transcript, and protein sequences. Rename the sequence files in the format of '**Gene name\_gene/rna/protein**' and keep the original file extension.

### B-H1 BarH1 [*Drosophila melanogaster* (fruit fly) ]

[Download Datasets](#)

Gene ID: 32724, updated on 23-Jun-2022

**Summary**

**Official Symbol** B-H1 provided by [FlyBase](#)

**Official Full Name** BarH1 provided by [FlyBase](#)

**Primary source** [FLYBASE:FBgn0011758](#)

**Locus tag** Dmel\_CG5529

**See related** [AllianceGenome:FB:FBgn0011758](#)

**Gene type** protein coding

**RefSeq status** REVIEWED

**Organism** [Drosophila melanogaster](#)

**Lineage** Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Holometabola; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora

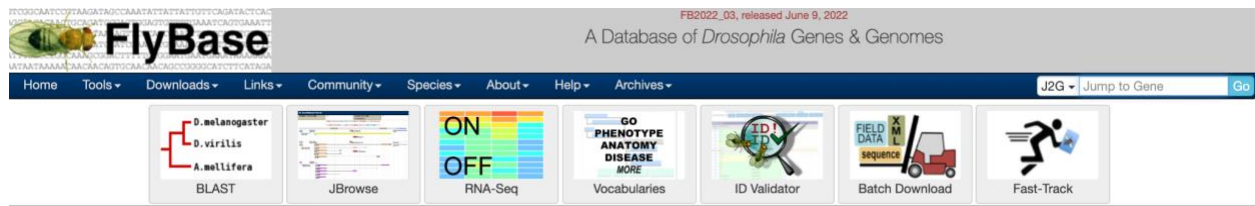
**Also known as** Bar; Bar H1; bar-h1; Bar-H1; BarH; barh1; barH1; Barh1; BarHI; BH1; CG5529; DmelCG5529

**Summary** Enables DNA-binding transcription factor activity, RNA polymerase II-specific. Involved in several processes, including animal organ morphogenesis; eye pigment granule organization; and leg disc proximal/distal pattern formation. Located in nucleus. Is expressed in several structures, including extended germ band embryo; larval segment; nervous system; presumptive embryonic/larval system; and sensory nervous system primordium. Human ortholog(s) of this gene implicated in oral squamous cell carcinoma. Orthologous to several human genes including BARHL2 (BarH like homeobox 2). [provided by Alliance of Genome Resources, Apr 2022]

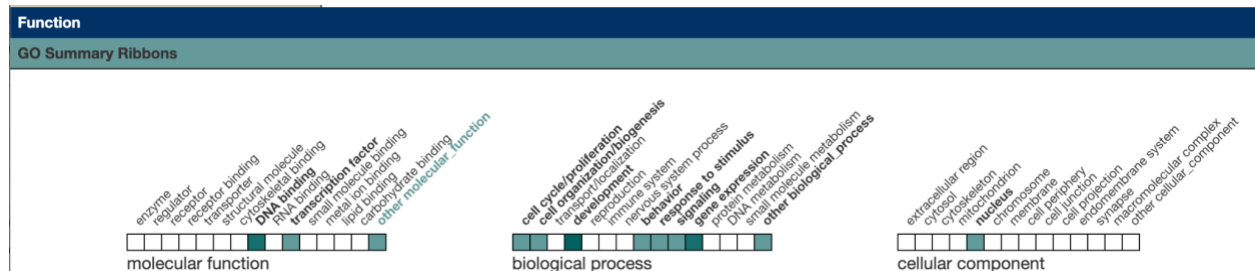
**NEW** Try the new [Gene table](#)  
Try the new [Transcript table](#)

- Read the hit report and fill up the table in the **result section** below.
- When looking for Gene Ontology (GO Summary), please go searching at Flybase as it provides more details.





- Search gene in **Flybase** starts by clicking on the dropbox at the top right corner of the page and then choosing 'Jump to Gene(J2G)'. After that, type the desired gene name or symbol in the search box and click on the search icon.
- Find the section of GO SUMmary Ribbons and pick the main (with the deepest color) molecular function, biological process, or cellular component.



## Assignment 5

- Lab Exercise II: Complete the table according to the Blast result.

Query sequence	Blast method	Species	Possible gene or protein name
Unknown gene sequence			
Unknown protein sequence			

- Lab Exercise III: Download these required genes and fill out the table below.

Gene name	Gene type	Gene Length	Exon count	Chromosome	GO-molecular function	GO-Biological process	GO-cellular component
Bar (B/BH1)							
Eyeless (ey)							

Vestigial (vg)							
Ebony (e)							

3. Create a word file, screenshot the sequence files you have downloaded with the correct name format and copy the above 2 tables to the word file.

## Lab 6 Gene structural annotation and sequence analysis

Hanqi Ye

### Objective

- Understand gene structure and be able to identify different gene elements
- To be familiar with alternative splicing and how to predict splice donor and acceptor sites in a gene
- Practice web-based gene sequence analysis tools: genome browsers and CD-search.

### Introduction

**Gene structure** is the organization of specialized sequence elements within a gene. Genes contain most of the information necessary for living cells to survive and reproduce. A gene is transcribed (copied) from DNA into RNA, which can either be non-coding (ncRNA) with a direct function, or an intermediate messenger (mRNA) that is then translated into protein. Each of these steps is controlled by specific sequence elements, or regions, within the gene. Every gene, therefore, requires multiple sequence elements to be functional (Figure 1).

Much of the gene structure is broadly similar between eukaryotes and prokaryotes. These common elements largely result from the shared ancestry of cellular life in organisms over 2 billion years ago. Understanding gene structure is the foundation of understanding gene annotation, expression, and function.

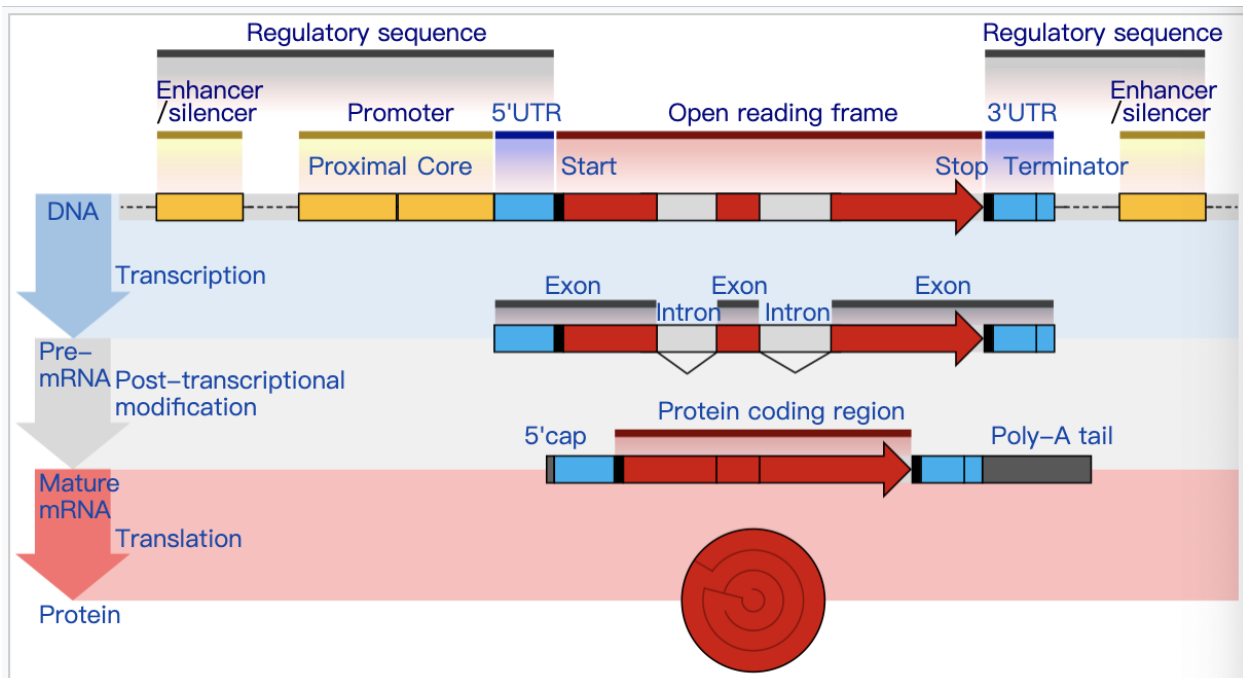


Figure 1. Annotated gene structure in eukaryotes. (Credit: wikipedia.org)

Common genomic features are listed below:

Regulatory sequence: a segment of a nucleic acid molecule that is capable of increasing or decreasing the expression of specific genes within an organism. Regulation of gene expression is an essential feature of all living organisms and viruses.

Open Reading Frame (ORF): spans of DNA sequence between the start and stop codons.

Promoter: a sequence of DNA to which proteins bind to initiate transcription of a single RNA transcript from the DNA downstream of the promoter.

Enhancer: a short (50–1500 bp) region of DNA that can be bound by proteins (activators) to increase the likelihood that transcription of a particular gene will occur.

Silencer: a sequence-specific element that induces a negative effect on the transcription of its particular gene. The most common position is found 20 bp to -2000 bp upstream of the target gene where it can help repress the transcription of the gene. Certain silencers can be found downstream of a promoter located within the intron or exon of the gene itself. Silencers have also been found within the 3' prime untranslated regions (3' UTR) of mRNA.

Terminator: a section of nucleic acid sequence that marks the end of a gene or operon in genomic DNA during transcription.

Untranslated region(UTR): an untranslated region (or UTR) refers to either of two sections, one on each side of a coding sequence on a strand of mRNA. If it is found on the 5' side, it is called the 5' UTR (or leader sequence), or if it is found on the 3' side, it is called the 3' UTR (or trailer sequence).

Exon: any part of a gene that will form a part of the final mature RNA produced by that gene after introns have been removed by RNA splicing

Intron: any nucleotide sequence within a gene that is removed by RNA processing during the production of the final RNA product.

Start: the first codon of a messenger RNA (mRNA) transcript translated by a ribosome.

Stop: a codon (nucleotide triplet within messenger RNA) that signals the termination of the translation process of the current protein.

Protein coding region: The coding region of a gene, also known as the coding sequence (CDS), is the portion of a gene's DNA or RNA that codes for protein.

Key differences in gene structure between eukaryotes and prokaryotes reflect their divergent transcription and translation machinery. For example, a key feature of the structure of eukaryotic genes is that their transcripts are typically subdivided into exon and intron regions.

Exon regions are retained in the final mature mRNA molecule, while intron regions are spliced out (excised) during post-transcriptional processing.

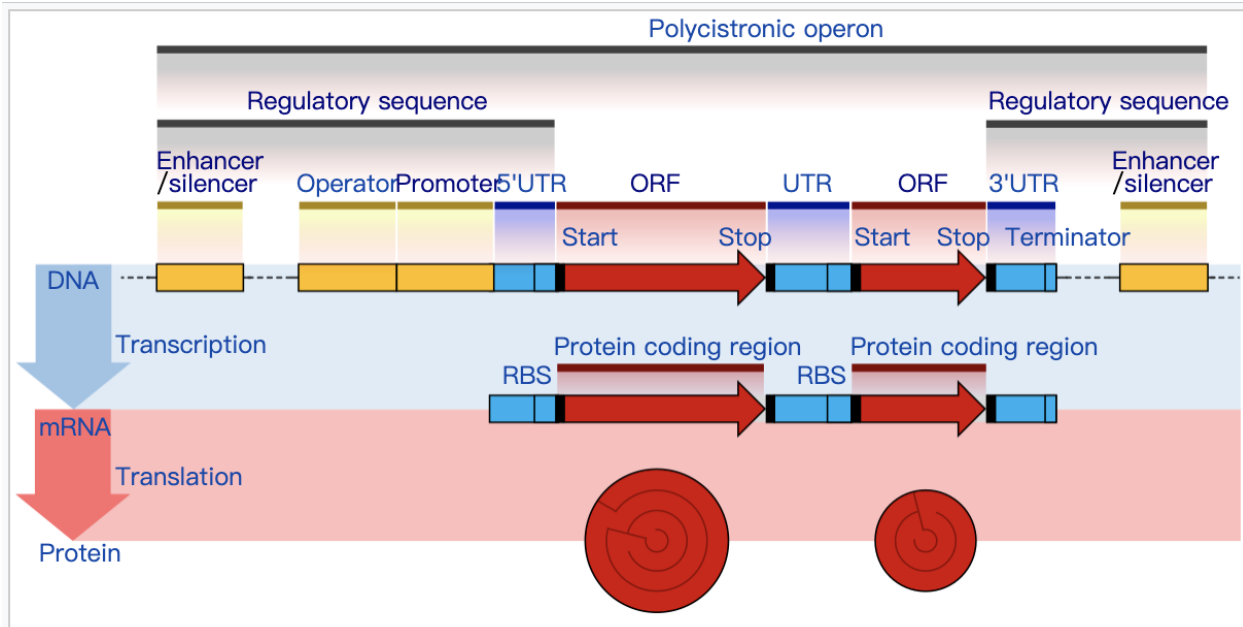


Figure 2. Annotated gene structure in Prokaryotes.(Credit: wikipedia.org)

Within introns, a splice donor site (5' end of the intron), a branch site (near the 3' end of the intron), and a splice acceptor site (3' end of the intron) are required for splicing. The splice donor site includes an almost invariant sequence **GU** (or **GT** when represented in DNA) at the 5' end of the intron, within a larger, less highly conserved region. The splice acceptor site at the 3' end of the intron terminates the intron with an almost invariant **AG** sequence. Upstream (5'-ward) from the AG there is a region high in pyrimidines (C and U), or polypyrimidine tract. Further upstream from the polypyrimidine tract is the branchpoint, which includes an adenine nucleotide involved in lariat formation. The consensus sequence for an intron (in IUPAC nucleic acid notation, see the table at right side) is: G-G-[cut]-G-U-R-A-G-U (donor site) ... intron sequence ... Y-U-R-A-C (branch sequence 20-50 nucleotides upstream of acceptor site) ... Y-rich-N-C-A-G-[cut]-G (acceptor site).

Description	Symbol	Bases represented					Complementary bases
		No.	A	C	G	T	
Adenine	A	1	A				T
Cytosine	C			C			G
Guanine	G				G		C
Thymine	T					T	A
Uracil	U					U	A
Weak	W	2	A			T	W
Strong	S			C	G		S
Amino	M		A	C			K
Ketone	K				G	T	M
Purine	R		A		G		Y
Pyrimidine	Y		C		T	R	
Not A	B	3		C	G	T	V
Not C	D		A		G	T	H
Not G	H		A	C		T	D
Not T <sup>a</sup>	V		A	C	G		B
Any one base	N	4	A	C	G	T	N
Gap	–	0					–

a. <sup>a</sup> Not U for RNA



Figure 3. Simple illustration of exons and introns in pre-mRNA (Credit: wikipedia.org)

Alternative splicing: (alternative RNA splicing or differential splicing) is an alternative splicing process during gene expression that allows a single gene to code for multiple proteins. In this process, particular exons of a gene may be included within or excluded from the final, processed mRNA produced from that gene. This means the exons are joined in different combinations, leading to different (alternative) mRNA strands. Consequently, the proteins translated from alternatively spliced mRNAs will contain differences in their amino acid sequence and, often, in their biological functions (see Figure3). This ensures that exons in the same gene are combined in different forms, so that a gene can produce different proteins at different times and in different environments, increasing the complexity or adaptability of the system under physiological conditions.

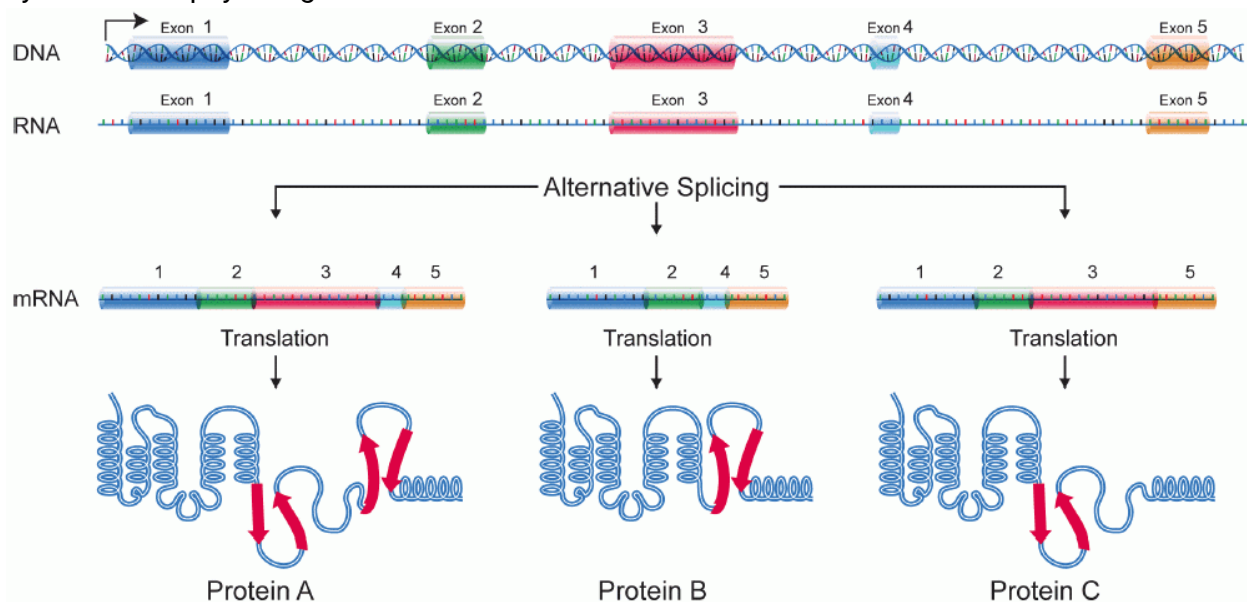


Figure 5. Alternative splicing produces three protein isoforms. Protein A includes all of the exons, whereas Proteins B and C result from exon skipping. (Credit: wikipedia.org)

There are 5 common alternative splice modes.

Exon skipping: in this case, an exon may be spliced out of the primary transcript or retained. This is the most common mode in mammalian pre-mRNAs.

Intron retention: A sequence may be spliced out as an intron or simply retained. This is distinguished from exon skipping because the retained sequence is not flanked by introns. If the

retained intron is in the coding region, the intron must encode amino acids in the frame with the neighboring exons, or a stop codon or a shift in the reading frame will cause the protein to be non-functional. This is the rarest mode in mammals but the most common in plants.

Mutually exclusive exons: One of two exons is retained in mRNAs after splicing, but not both.

Alternative donor site: An alternative 5' splice junction (donor site) is used, changing the 3' boundary of the upstream exon.

Alternative acceptor site: An alternative 3' splice junction (acceptor site) is used, changing the 5' boundary of the downstream exon.

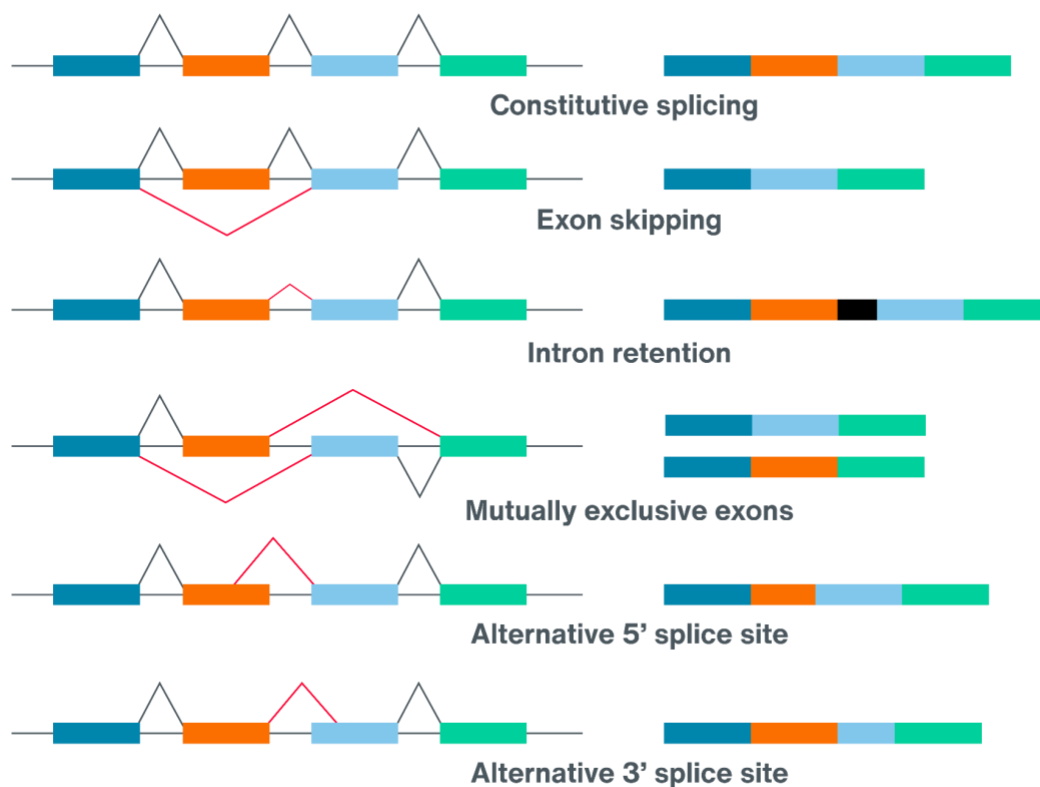


Figure 6. Constitutive splicing with five basic modes of alternative splicing. (Credit: nanoporetech.com)

### Lab Exercise I - Identify gene elements.

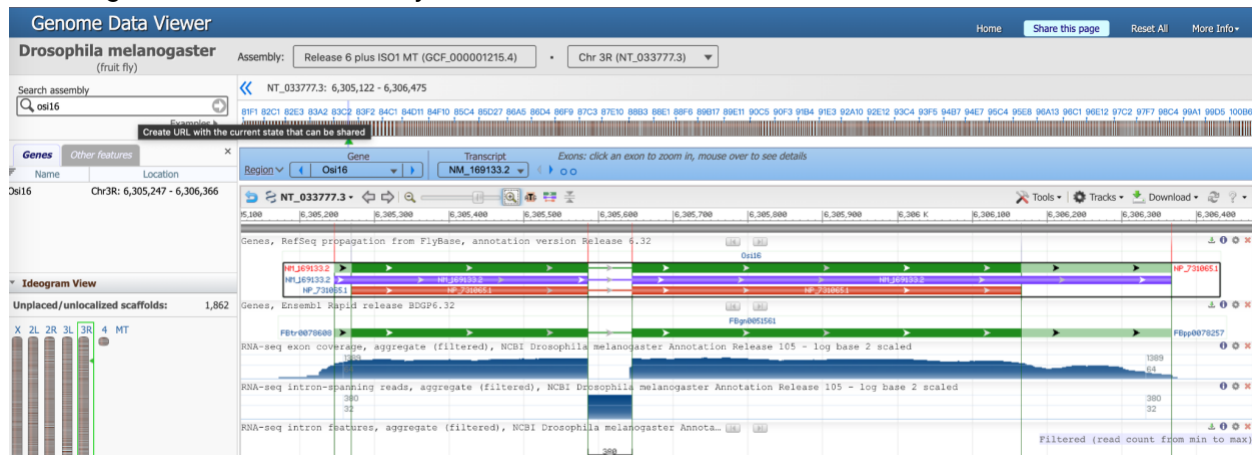
#### Material

Laptop

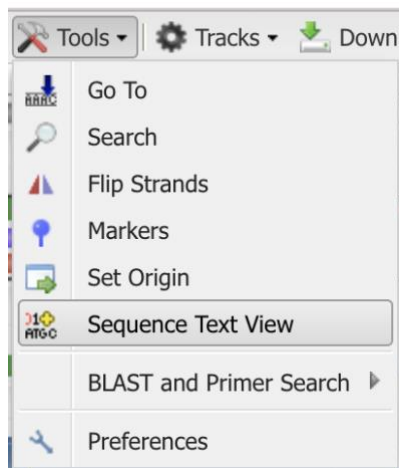
[NCBI genome data viewer](#)

## Methods

1. Use the gene Osi16 as an example. Open NCBI genome data viewer and search for the gene Osi16 in the fruit fly.



2. Click on the Tool icon and choose the Sequence text view



3. Once the text view is ready, move your mouse to locate the gene Osi16. Purple highlighted sequences are UTRs. The Green highlighted sequence is the intron. Pink sequences are exon.

```

CGAGATGATGCAGACTCACAAACAGTCTCCATTGCAGTAGATAGCGAAGATGTTGCTGACCAAGACCGTGAAATATCTTTTTTACTTTGGCCTTATTCGCC
M L L T K T V K Y L F Y L A L F A
TTCATGTGCAAAATATGCCACGGCTGCCAGTGTCCAGCCCACCGAAGTGCCTGCAGTTGCTCCGGAACCATAGAAATACCCAAAGGGCAGAGAGCCTTC
F M C K Y A T A A S V Q P T E V P A V A P E T I R I P Q R A E S L
TCAGTGGCTGTGAAGCATCTTCCTTCAGTTGGATGTGCTCAAGATCGAGTTTGTCAAGATCATGAAAAGCTTGCCGAAACAGGAGGAACAAACGTTTT
L S G C E A S S F S W M C L K I E F V K I M E K L A E Q E E L N V L
GCCAGGAATCAGCGTGGTCAAGGATGAGAACGCCACCGAACTCAAGACATCGGAGCTCATGGCAGGTATGTTCTTGAAAACACAAAAATAAGACAAAA
P G I S V V K D E N A T E L K T S E L M A
ATTTTAACTTTCGACAATATACAGAAGTTGCTCGCAGCTATCCCAAGTGACCCGAGCACTCGCTTGAATGGCTATATAGTAGCCAAACTGGAAAACCTTCT
E V A R S Y P S D P S T R L N G Y I V A K L E N L L
GAGAACCGGTTCTCTACGATTTCCGCTGCTGGATGCAAGTCCCTCGTGAAGGTCGCAAGCACAAAGTTCGGCAAGAAAGGAGGCTTGGAGGCCCTAGTG
R T R F L D F L D K S L V E G R K H K F G K K G L E A L V
GCGGCTGGAGTCATGATGAAGGGAATGCTTATGGCCATGGGCTCGGGGCCATCGCTCTGATGGCCGAAAGGCTCTGATGACAGCTTTAATGGCCTTGA
A A G V M M K G M L M A M G L G A I A L M A G K A L M T A L M A L
CCCTGTCTGGAGTCCCTCGGTCTAAAGAGCTTGGCGGTGGCGGGGAAAGTCCACACATACGAAATAGTGGCTAAACCAATTTACACCTCCAGCCACTC
T L S G V L G L K S L A G G G K S T T Y E I V A K P I Y T S S H S
GCACCTCGGTACGACGAAAGATGGGAGGTCACTCCACAGTCCCACTTTGCCGCAAGGAGGTGGGGGTGGTACTGCCAGTGGCTTTGGCTACGGAGGA
H S V T H E D G G H S H S P H F A A G G G G G G T A S G F G Y G G
TATGCGAGGTCCCTGAAGGATCAATCAGCGAACAATACTAATAATTTAGAAAAGAACCAGACAGTATCATTACCTAATAACTAATTAATTTATGTT
Y A R S L K V D Q S A N K I *
ACTTTAAGAAACACAATATCTCAGCAAACCTCAGCATTTGTTTTTTGTTTTATTGTGTGAACAGAACAGTTAAATAGACGTTGCCGATCTTCCAAATAATCATTT
AGTCAGATTTTATTACTTTTAAATATGCCCAAGGTTTTTTATAAACGACTATTAAAGCCACAAACGTAAAAAAATGTTTAAATTTTATTTTGAATAAAT

```



4. Screenshot the whole gene text to assignment 6 report and put sequence elements (5' UTR, 3' UTR, Exons, Introns, start codon by a red square, stop by a blue square, splice donor site by red cycle, and splice acceptor by blue cycle ) labels on it.

5' UTR

CGAGATGATGCAGACTCACAAACAGTCTCCATTGCAGTAGATAGCGAAGATGCTGCTGACCAAGACCGTGAAATATCTTTTTTACTTGGCCCTTATTCGCC

Exon1

Exon2

Intron

3' UTR

AGTCAGATTTTATTACTTTTAAATATGCCAAGCTTTTATAAACGACTATTAAGCCACAAACGTAAAAAATGTTTAAATTTTTATTGGAATAAT

## Lab Exercise II - Identify alternative splicing mode of a gene and explore RNA expression in Flybase Jbrowser.

### Material

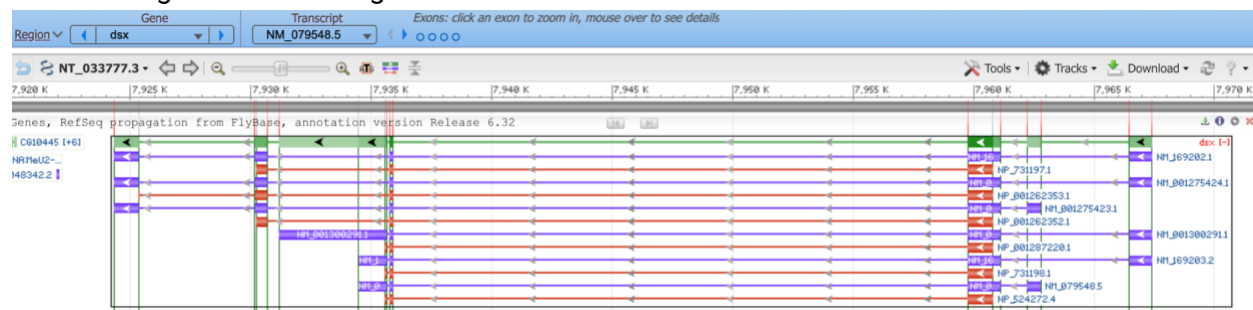
Laptop

[NCBI genome data viewer](#)

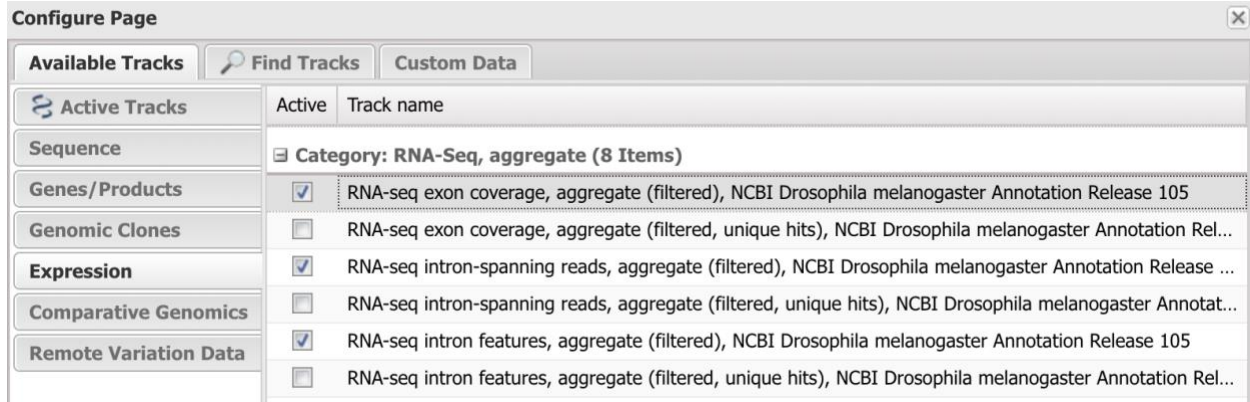
[Flybase Jbrowser](#)

### Methods

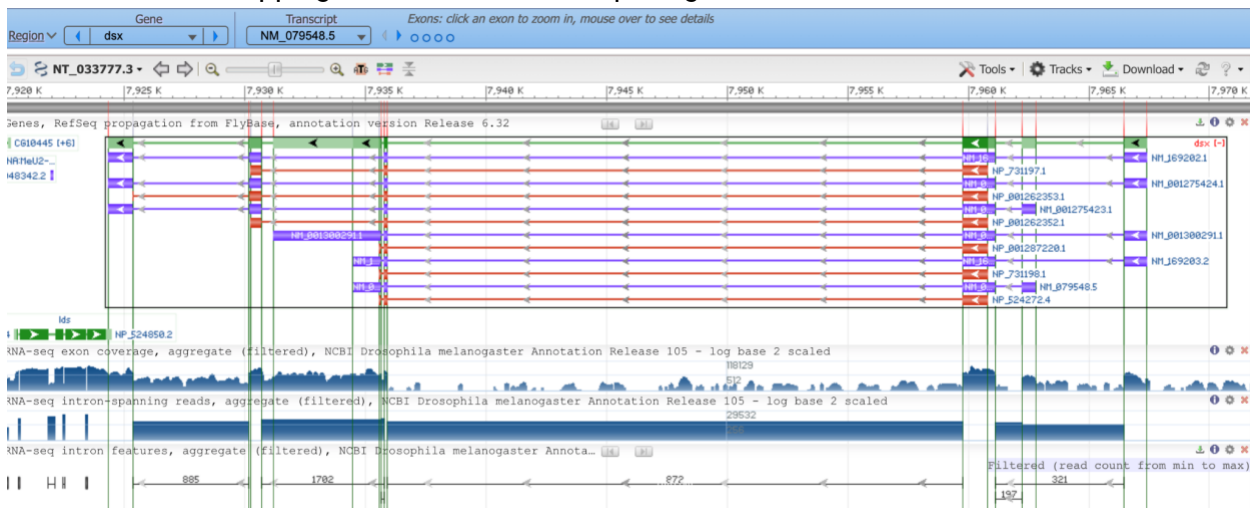
1. Take gene doublesex as an example. Search dsx in NCBI genome data viewer. Expand the green bar of the gene.



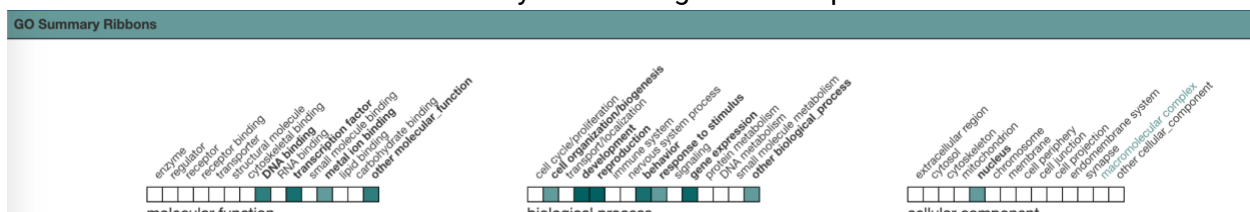
2. Click on the Tracks icon and choose 'Configure tracks'. Make sure you choose RNA-seq exon coverage, RNA-seq intron-spanning reads, and RNA-seq intron features.



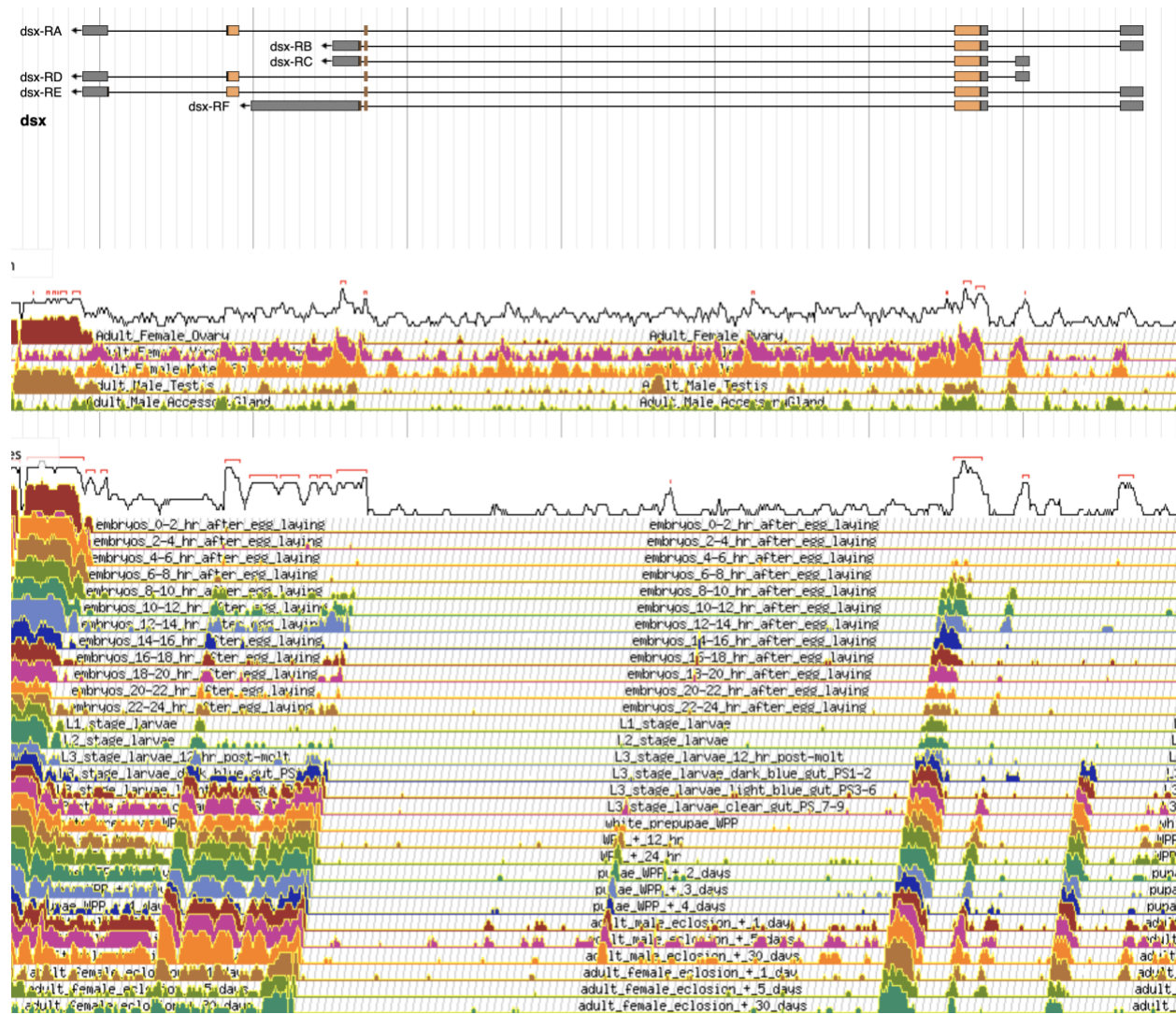
- Save the screenshot below to the assignment 6 report. In the report, determine which alternative splicing mode occurs in each isoform of mRNAs (ID starts with NM) compared with the green constitutive splicing. For example, the *D. melanogaster* gene *dsx* contains 7 exons and reported 6 isoforms of mRNA. Isoform NM\_169202.the 1 is the exon skipping mode of alternative splicing since it has no exon 6.



- Still use the *dsx* gene as an example. Go to Jbrowser and search for the *dsx* gene. Take a screenshot of the GO summary to the assignment 6 report.



- In the left navigation, you can add more tracks to the viewer. According to the GO summary, the *dsx* gene is involved in development and reproduction, we can find it under the Expression section and choose Reproductive system and development stages RNA expression tracks. Screenshot result to the assignment.



- In assignment 6, write a summary about this gene according to the RNA expression result. For example, the *dsx* gene shows the differential RNA expression pattern between males and females, which matches its function of regulating sexual development.

### Lab Exercise III - Comparative genomics in UCSC genome browser and protein conservation analysis in CD-search NCBI tool.

#### Material

Laptop

Target gene's protein sequence file downloaded in lab 5.

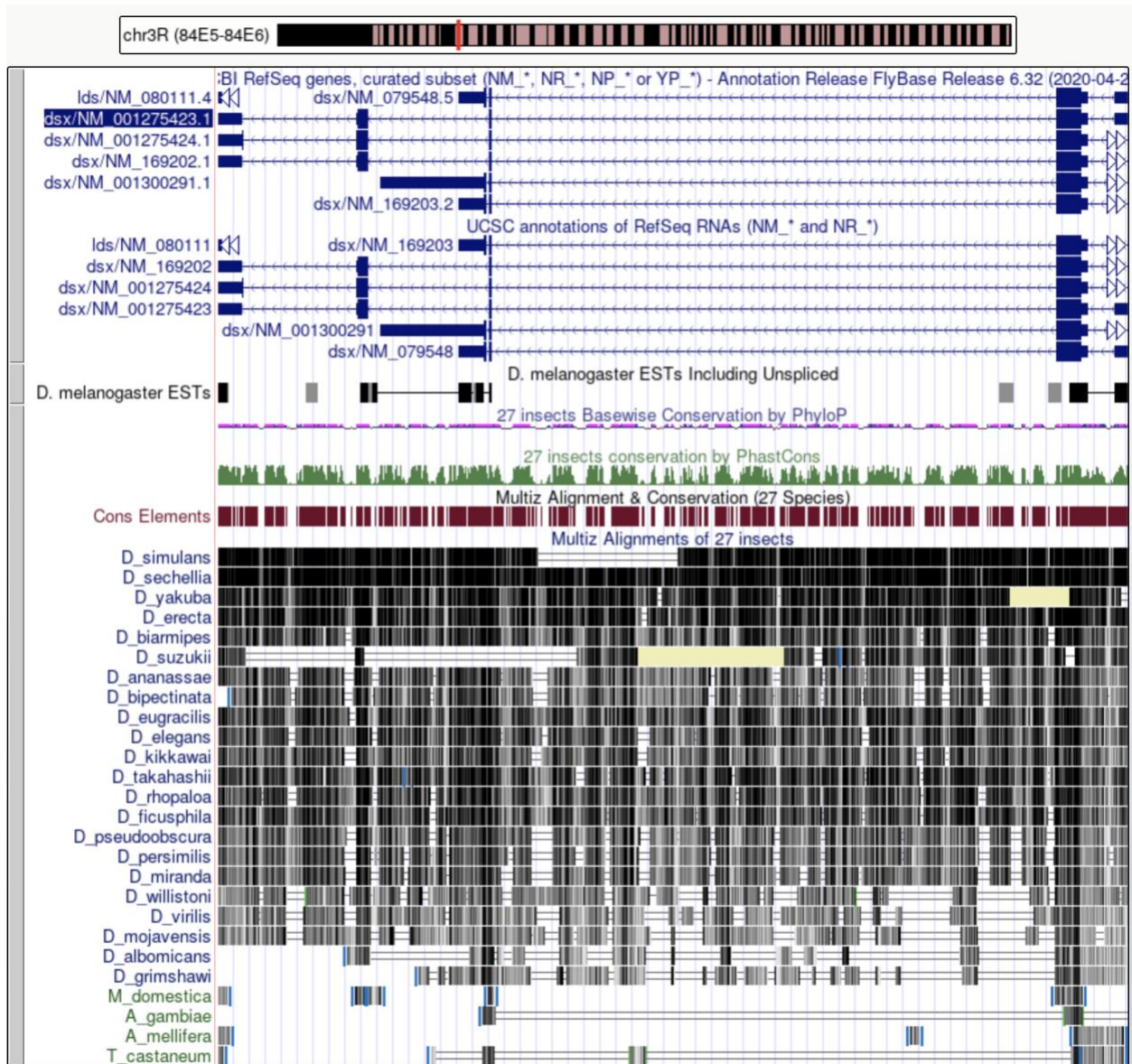
[UCSC genome browser](#)

[NCBI Conserved Domains search \(CD-search\)](#)

#### Methods



- For comparative genomics, please go to the UCSC genome browser and search the dsx gene. Scroll down the page and find track options. Choose only NCBI Refseq, D. melanogaster ESTs, and Conservation (pack) tracks. Save the screenshot to assignment 6.



- Open the NCBI CD-search tool to find the conserved protein domain in a protein sequence. Find the *dsx* protein sequence file, and copy and paste only one sequence to the entry box. Use the default parameters set in the options and click on Submit.

NCBI

HOME | SEARCH | GUIDE | Structure Home | 3D Macromolecular Structures | Conserved Domains

Search for **Conserved Domains** within a protein or coding nucleotide sequence

Enter **protein** or **nucleotide** query as accession, gi, or sequence in **FASTA format**. For multiple protein queries, use **Batch CD-Search**.

Submit Reset Help

**OPTIONS**

Search against database:

Expect Value threshold:

Apply low-complexity filter: ☐

Composition based statistics adjustment: ☒

Force live search: ☐

Rescue borderline hits: ☐ Suppress weak overlapping hits: ☐

Maximum number of hits:

Result mode: ☒ Concise ☐ Standard ☐ Full

- Here is the result, expand all the domain hits. Screenshot the result and save to the assignment 6 report.

Conserved domains on [lcl|seqsig\_MVSEE\_991785ce39c4f50b28271b0bf2189f88] View

NP\_001262352.1 dsx [organism=Drosophila melanogaster] [GeneID=40940] [isoform=D]

**Protein Classification**

**DM and DSX\_dimer domain-containing protein** (domain architecture ID 10465243)  
DM and DSX\_dimer domain-containing protein

**Graphical summary** ☐ Zoom to residue level show extra options

Query seq. Specific hits Superfamilies

DM DSX\_dimer

Search for similar domain architectures Refine search

**List of domain hits**

Name	Accession	Description	Interval	E-value
DM	pfam00751	DM DNA binding domain; The DM domain is named after dsx and mab-3. dsx contains a single ...	41-86	3.08e-26
DM DNA binding domain; The DM domain is named after dsx and mab-3. dsx contains a single amino-terminal DM domain, whereas mab-3 contains two amino-terminal domains. The DM domain has a pattern of conserved zinc chelating residues C2H2C4. The dsx DM domain has been shown to dimerize and bind palindromic DNA.				
Pssm-ID: 395608 Cd Length: 47 Bit Score: 100.82 E-value: 3.08e-26				
seqsig_MVSEE_991785ce39c4f50b28271b0bf2189f88 41 PPNCARCRNHGLKITLKGHKRYCKFRYCTCEKCRLTADRQVRVMAQ 86 Cdd:pfam00751 2 SPKCARCRNHGVLPVLPKGHKRYCPYRDCTCPKCLLVAERQVRVMAQ 47				
DSX_dimer super family	sf07436	Doublesex dimerization domain; Doublesex (DSX) is a transcription factor that regulates ...	354-403	9.34e-16
Doublesex dimerization domain; Doublesex (DSX) is a transcription factor that regulates somatic sexual differences in Drosophila. The structure of this domain has revealed a novel dimeric arrangement of ubiquitin-associated folds that has not previously been identified in a transcription factor.				
The actual alignment was detected with superfamily member pfam08828:				
Pssm-ID: 370143 Cd Length: 60 Bit Score: 71.64 E-value: 9.34e-16				
seqsig_MVSEE_991785ce39c4f50b28271b0bf2189f88 354 DVFLDYQKLEKFRYPWELMPLMYVILKDADANIEEASRIIEARVEIN 403 Cdd:pfam08828 2 DELVKRAQWLEKLGYPWELMPLMYVILKSDAGDVQKARQIDEDQAEIR 51				

- Write a summary about the conservation of the gene. For example, the dsx gene is more conservative in D\_sechellia than other insects since it has less gap and a darker color (darker means more conservative a gene is) as shown in the UCSC genome browser. In addition, the dsx conserved domain DM and DSX\_dimer locations match the conversation signals in the UCSC genome browser Conservation track.

## Assignment 6

- Create a word file as an assignment 6 report.

2. Do Lab exercise I with gene Bar, and search for B-H1. B-H1 gene has 2 isoforms of protein, choose the longer one for labeling and save the result to the assignment 6 report.
3. Do Lab exercise II with gene eyeless, save screenshot and summary to assignment 6.
4. Do Lab exercise III with gene Ebony, save screenshot and summary to assignment 6.

## **Lab 7 PCR Primer design, download, and validation**

Hanqi

### **Objective**

- To be familiar with the principle of designing a primer
- Practice primer design and validation via NCBI primer blast tool
- Get familiar with FlyPrimerBank.

### **Introduction**

PCR is a standard technology in biological laboratories. Biotechniques like gene quantification, cloning, assembly, mutation, and sequencing cannot perform without PCR. At present, PCR-related reagents, consumables and equipment have been well developed, including the optimization of polymerase and buffer, which has greatly improved the amplification capability, fidelity, and compatibility. Therefore, we no longer need to distract some of our attention, as in the past, to explore the thermal cycle conditions, adjust the concentration of the substrate of the reaction system, etc. In other words, the whole PCR experimental conditions can be standardized, without having to touch the conditions from scratch for each experiment. However, in the PCR experiment, one step that the manufacturer cannot help us to optimize is primer design.

PCR primers are short pieces of single-stranded DNA, usually around 20 nucleotides in length. Two primers are used in each PCR reaction, and they are designed so that they flank the target region (a region that should be copied). That is, they are given sequences that will make them bind to opposite strands of the template DNA, just at the edges of the region to be copied. The primers bind to the template by complementary base pairing rules. When the primers are bound to the template, they can be extended by the polymerase, and the region that lies between them will get copied.

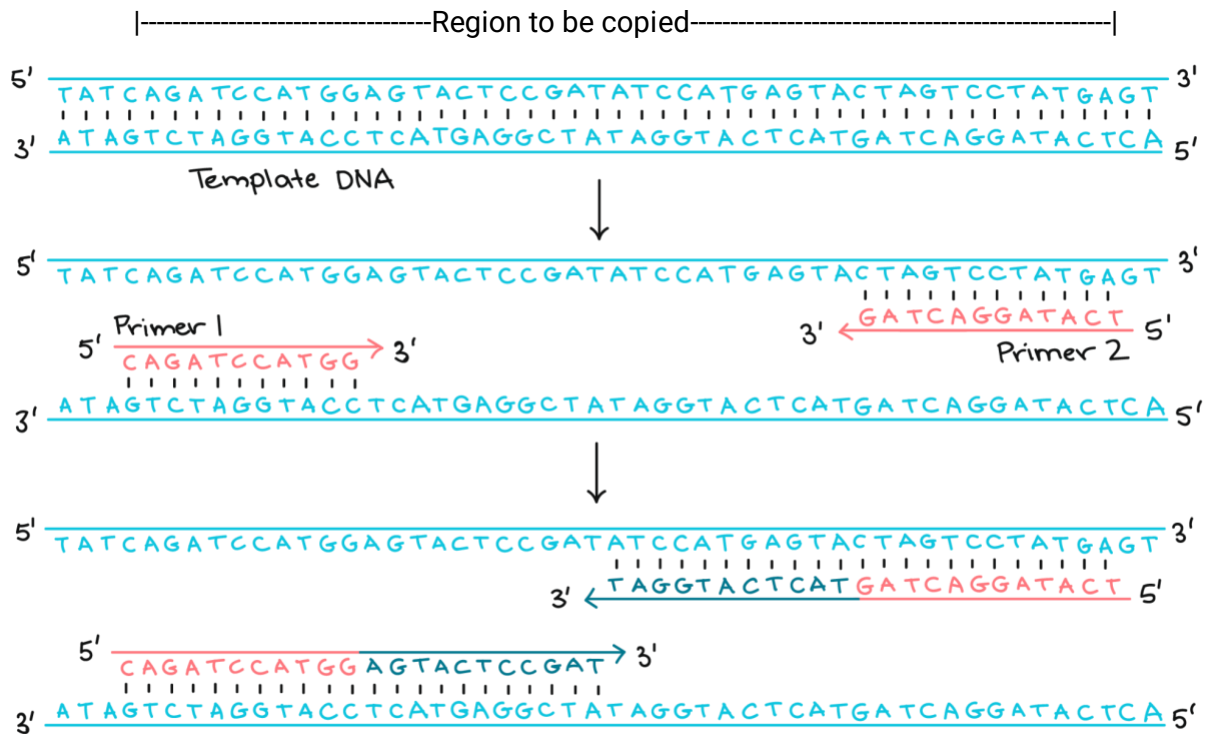


Figure 1. Primers bind to template DNA in the flanking region in PCR cycles. (Credit: Khanacademy.org)

Thus, different target genes require a specific pair of primers. The first exercise of this lab will take one gene (eyeless) as an example and teach you to design primers for PCR experiments by using the primer blast tool in NCBI. In exercise 2, we will briefly describe how to validate primers using the same tool. Exercise 3, practice downloading primers from FlyprimerBank and validating them.

### Lab Exercise I - PCR Primer design.

#### Material

Student needs to bring their own laptop

NCBI website (<https://www.ncbi.nlm.nih.gov/>)

NCBI Primer-Blast tool (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi>)

#### Methods

##### **A) Design a primer for a specific transcript.**

1. Open NCBI and search for the gene eyeless(ey). Choose the gene and scroll down the result page, find the genomic viewer, and click on the ey [+8] the green bar shown below.



Genomic regions, transcripts, and products

Go to [reference sequence details](#)

Genomic Sequence: NC\_004353.4 Chromosome 4 Reference Release 6 plus ISO1 MT Primary Assembly

Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)

- More information will show up including four transcript variants for this gene and four protein variants. (NM is mRNA, NP is protein, NC is a nucleotide.)

- Continue scrolling down and find **NCBI Reference Sequences (RefSeq)**, **mRNA**, and **protein** section. specific information about all isoforms will be displayed if there is more than one mRNA transcript. In this case, is A to D eyeless transcripts.

mRNA and Protein(s)

- [NM\\_166789.2](#) → [NP\\_726607.1](#) **eyeless, isoform B [Drosophila melanogaster]**  
[See identical proteins and their annotated locations for NP\\_726607.1](#)  
 Status: REVIEWED

- Using isoform B as an example, this primer is designed specifically for amplifying eyeless isoform B cDNA during the PCR. click on the hyperlink [NM\\_166789.2](#) to go to the following page.

#### Drosophila melanogaster eyeless (ey), transcript variant B, mRNA

NCBI Reference Sequence: NM\_166789.2

[FASTA](#) [Graphics](#)

Go to: ☐

LOCUS NM\_166789 2854 bp mRNA linear INV 22-APR-2020  
 DEFINITION Drosophila melanogaster eyeless (ey), transcript variant B, mRNA.  
 ACCESSION NM\_166789  
 VERSION NM\_166789.2  
 DBLINK BioProject: [PRJNA164](#)  
 BioSample: [SAMN02803731](#)  
 KEYWORDS RefSeq...

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Show in Genome Data Viewer

- Click the Pick Primers on the right side under customize the view to enter the Primer Blast page. Once you see the page, PCR Template has been filled out for you. For now, leave everything else blank and change the Max of **PCR product size** to 350. In qPCR reactions, the extension time generally does not exceed 30 SEC. According to the conventional Taq enzyme reaction rate of 1 KB /min, 30 SEC can synthesize 500 bp.

However, the enzyme activity will decrease with the prolongation of the reaction time, so it is more appropriate to set the maximum product length as 350 bp or even lower to be conservative. Too short is not good because it can't be separated from the primer dimer, so the lower limit is 70. If you can't find any suitable primers, then increase this number, preferably below 500, or else the qPCR reaction conditions will have to be changed (and therefore the extension time).

**Primer-BLAST**
A tool for finding specific primers

Finding primers specific to your PCR template (using Primer3 and BLAST).

Primers for target on one template

Primers common for a group of sequences

Retrieve recent results
Publication
Tips for finding specific primers

**PCR Template**

Enter accession, gi, or FASTA sequence (A refseq record is preferred) ? Clear

Or, upload FASTA file 选择文件 未选择任何文件

Range ? Clear

	From	To
Forward primer	<div style="border: 1px solid #add8e6; width: 50px; height: 20px; margin: 2px;"></div>	<div style="border: 1px solid #add8e6; width: 50px; height: 20px; margin: 2px;"></div>
Reverse primer	<div style="border: 1px solid #add8e6; width: 50px; height: 20px; margin: 2px;"></div>	<div style="border: 1px solid #add8e6; width: 50px; height: 20px; margin: 2px;"></div>

**Primer Parameters**

Use my own forward primer (5'→3' on plus strand) ? Clear

Use my own reverse primer (5'→3' on minus strand) ? Clear

PCR product size

Min	Max
<div style="border: 1px solid #add8e6; width: 40px; height: 20px; margin: 2px;"></div>	<div style="border: 1px solid #add8e6; width: 40px; height: 20px; margin: 2px;"></div>

# of primers to return

Primer melting temperatures (T<sub>m</sub>)

Min	Opt	Max	Max T <sub>m</sub> difference
<div style="border: 1px solid #add8e6; width: 40px; height: 20px; margin: 2px;"></div>	<div style="border: 1px solid #add8e6; width: 40px; height: 20px; margin: 2px;"></div>	<div style="border: 1px solid #add8e6; width: 40px; height: 20px; margin: 2px;"></div>	<div style="border: 1px solid #add8e6; width: 40px; height: 20px; margin: 2px;"></div> <span style="color: #007bff;">?</span>

6. The **Exon junction span** option is changed to 'Primer must span an exon-exon junction. This means that at least one of the primers crosses the junction between the exons. In this way, even if there is genomic DNA contaminating the cDNA, it will not be amplified, ensuring accurate quantification. Of course, if some genes simply do not have introns, changing this option will result in an error, thus go choose another option.

**Exon/intron selection**

Exon junction span

Exon junction match

Intron inclusion

Intron length range

A refseq mRNA sequence as PCR template input is required for options in the section ?

✓ No preference ?

Primer must span an exon-exon junction

Primer may not span an exon-exon junction

Minimal and maximal number of bases that must anneal to exons at the 5' or 3' side of the junction ?

☐ Primer pair must be separated by at least one intron on the corresponding genomic DNA ?

Min	Max
<div style="border: 1px solid #add8e6; width: 40px; height: 20px; margin: 2px;"></div>	<div style="border: 1px solid #add8e6; width: 40px; height: 20px; margin: 2px;"></div> <span style="color: #007bff;">?</span>

7. The following options are set as default. Specificity Check is chosen as default, which helps you verify the Specificity of a primer. The organism is automatically filled in for you *D.mel* (7227 is for fruit fly D.mel, add more species if you need.) The last option 'Allow splice variants' is not chosen for designing a specific primer for this chosen isoform.

42

**Primer Pair Specificity Checking Parameters**

**Specificity check**  
☒ Enable search for primer pairs specific to the intended PCR template ?

**Search mode**  
 Automatic ?

**Database**  
 Refseq mRNA ?

**Exclusion**  
☐ Exclude predicted Refseq transcripts (accession with XM, XR prefix) ☐ Exclude uncultured/environmental sample sequences ?

**Organism**  
 7227 [Add organism](#)  
 Enter an organism name (or organism group name such as enterobacteriaceae, rodents), taxonomy id or select from the suggestion list as you type. ?

**Entrez query (optional)**  
 ?

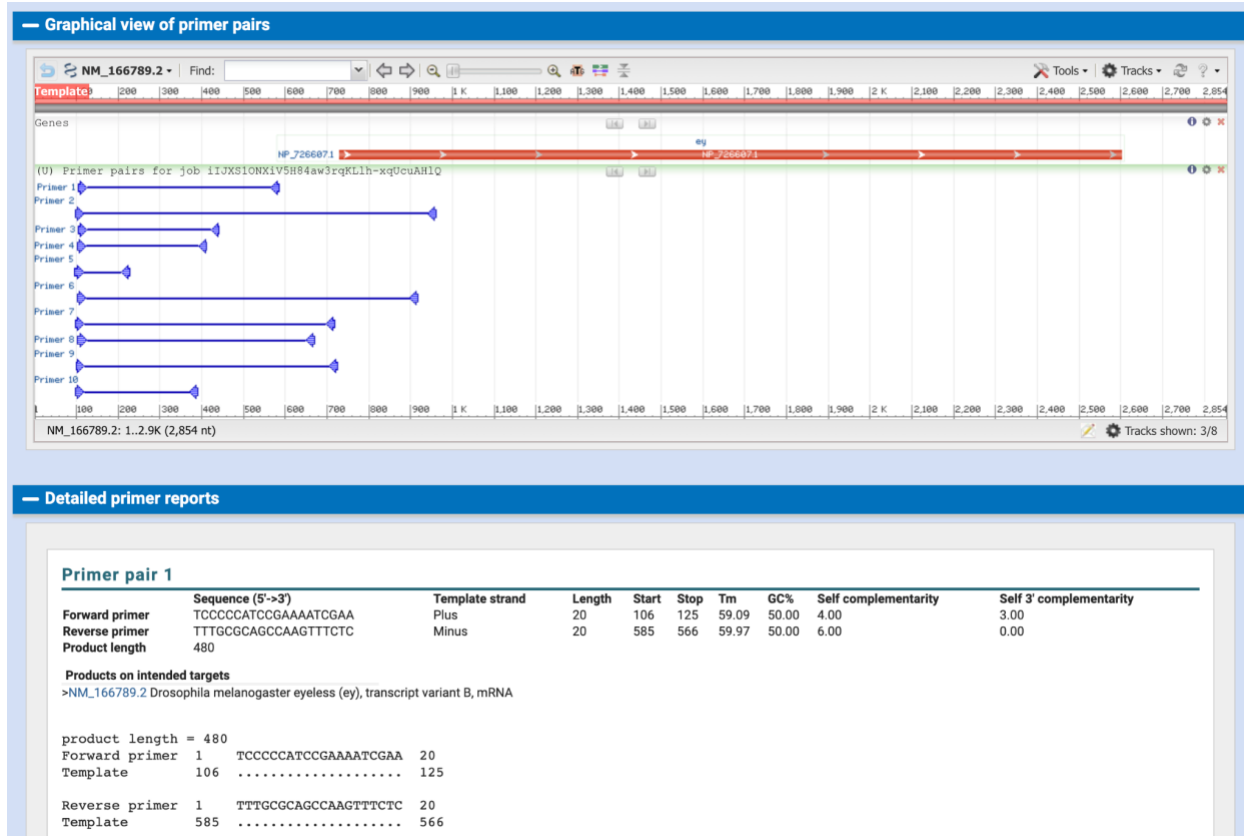
**Primer specificity stringency**  
 Primer must have at least 2 total mismatches to unintended targets, including at least 2 mismatches within the last 5 bps at the 3' end. ?  
 Ignore targets that have 6 or more mismatches to the primer. ?

**Max target amplicon size**  
 4000 ?

**Allow splice variants**  
☐ Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input) ?

[Get Primers](#) ☐ Show results in a new window ☒ Use new graphic view ?

8. Click the Get Primers button and wait for a few seconds or minutes depending on the number of search requests. Once done, this page returns 10 pairs of primers.

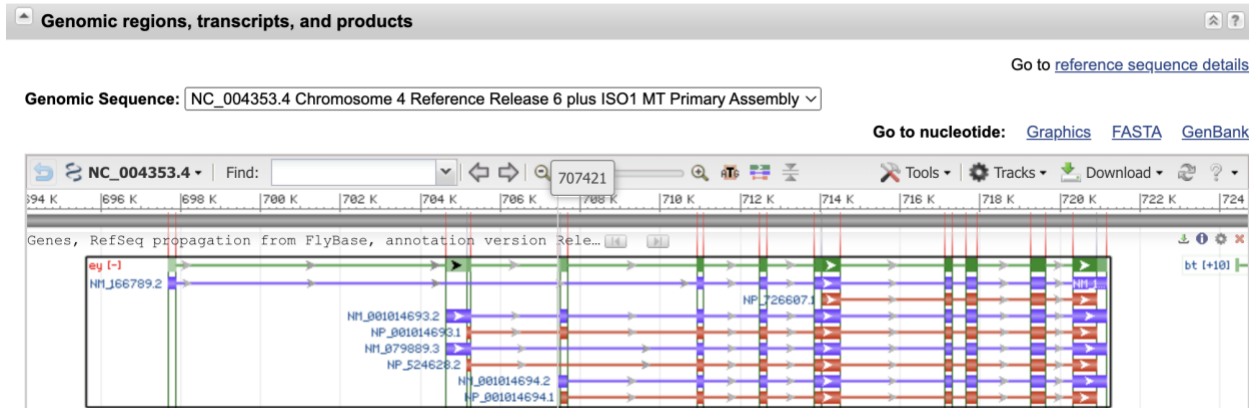


9. Save the screenshot of the Graphical view of primer pairs and copy the primer pair 1 text to your word report as well. **Be aware to indicate which gene this primer is designed for.**

## B) Design a primer for multiple isoforms

you need to change the range of the forward or reverse primer. The goal is to make sure that the primer falls in the common interval of all the transcripts since these isoforms have strong homology, basically only the 5' or 3' end is slightly different.

1. Find the common interval of all transcripts in the genomic regions viewer at the top of the gene page. In this case, the 3' end in all isoforms are the same, only the 5' end is different. We will choose the longest transcript isoform (NM\_166789.2) as the PCR template but use the range of the shortest transcript isoform (NM\_001014694.2).



2. Compare with other isoforms, NM\_001014694.2 is missing 2 exons. Move the mouse to the red lines indicating the range of exons to calculate the common range of all transcript variants. In this case, the first exon range is (697673-697874), 2nd the exon range is (704621-705257).
3. Click on the hyperlink of the longest isoform which is isoform B [NM\\_166789.2](#) and choose pick primers. In the Forward primer range, it begins from 873.

## Primers for target on one template

## Primers common for a group of sequences

### PCR Template

Retrieve recent results Publication Tips for finding specific primers

Enter accession, gi, or FASTA sequence (A refseq record is preferred) ?

Clear

NM\_166789.2

Range ?

Clear

Forward primer

From

To

Reverse primer

- Other settings are the same as the single isoform primer design except for choosing the **Allow splice variants**. Click on Submit and wait until you see a warning.

Allow splice variants

☒ Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input) ?

- The following warning is asking about which variant you want to include together for this primer. Choose the ones you want and click Submit again.

### Primer-BLAST

A tool for finding specific primers

Finding primers specific to your PCR template (using Primer3 and BLAST)

Input PCR template NM\_166789.2 Drosophila melanogaster eyeless (ey), transcript variant B, mRNA

Range 873 - 2854

Your PCR template is highly similar to the following sequence(s) from the search database. To increase the chance of finding specific primers, please review the list below and select all sequences (within the given sequence ranges) that are intended or allowed targets.

Select: All None Selected:3

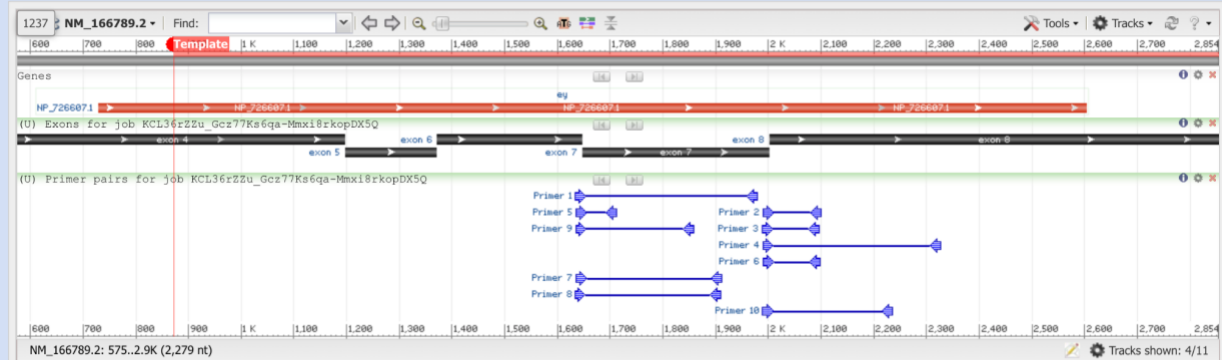
Accession	Title	Identity	Alignment length	Seq. start	Seq. stop
<input checked="" type="checkbox"/> NM_001014693.2	Drosophila melanogaster eyeless (ey), transcript variant D, mRNA	100%	1982	1481	3462
<input checked="" type="checkbox"/> NM_001014694.2	Drosophila melanogaster eyeless (ey), transcript variant C, mRNA	100%	1982	921	2902
<input checked="" type="checkbox"/> NM_079889.3	Drosophila melanogaster eyeless (ey), transcript variant A, mRNA	100%	1982	1301	3282

Submit

☒ Show results in a new window

- Wait until you see the result with 10 pairs of candidate primers.

### Graphical view of primer pairs



### Detailed primer reports

#### Primer pair 1

	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GGGTCCCTCAGTCAGTACCA	Plus	20	1634	1653	60.25	60.00	5.00	1.00
Reverse primer	CGCGTATGATGATGTTGG	Minus	20	1979	1960	60.04	55.00	6.00	0.00
Product length	346								
Exon junction	1647/1648 (forward primer) on template NM_166789.2								

- Save the screenshot of the Graphical view of primer pairs and copy the primer pair 1 text to your word report as well. **Be aware to indicate which isoforms this primer is designed for.**

## Lab Exercise II - Previous Primers validation

### Material

Laptop

Previous Primers given by instructor

NCBI Pick Primers tool (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi>)

### Methods

Let's use an eyeless transcript primer as an example.

- Pick your desired isoform, click on its NM hyperlink and then click on the pick primers.

#### **mRNA and Protein(s)**

- [NM\\_166789.2](#) → [NP\\_726607.1](#) **eyeless, isoform B [Drosophila melanogaster]**

[See identical proteins and their annotated locations for NP\\_726607.1](#)

- In the pick primers page, input the candidate forward primer and reverse primer for validation. Also, choose Exon junction span as Primer must span an exon-exon junction.

Primers for target on one template

Primers common for a group of sequences

Retrieve recent results

Publication

Tips for finding specific primers

PCR Template

Enter accession, gi, or FASTA sequence (A refseq record is preferred) ? Clear

NM\_166789.2

Or, upload FASTA file 选择文件 未选择任何文件

Range ? Clear

Forward primer From To

Reverse primer From To

Primer Parameters

Use my own forward primer (5'→3' on plus strand) ? Clear

Use my own reverse primer (5'→3' on minus strand) ? Clear

Min

Max

70

1000

PCR product size

# of primers to return 10

Min

Opt

Max

Max T<sub>m</sub> difference

57.0

60.0

63.0

3 ?

Primer melting temperatures (T<sub>m</sub>)

Exon/intron selection

A refseq mRNA sequence as PCR template input is required for options in the section ?

Exon junction span Primer must span an exon-exon junction ?

- If you see this warning, you can loosen the selection by unchecking the exon junction span and redo it.

**Primer-BLAST** » JOB ID: J5LV1SWWT5-BMkBXGHtM756\_AGTaeckg

Primer-BLAST Results

**Warnings:**  
 No primers satisfying exon/intron selection parameters were found. Try to loosen the selection criteria and/or increase value for "Max primer pairs to screen" parameter.

Input PCR template NM\_166789.2 Drosophila melanogaster eyeless (ey), transcript variant B, mRNA  
 Range 1 - 2854  
 Other reports [Search Summary](#)

No primers were found...see explanation below: Primer3 info:  
 Left primer: considered 1, high end compl 1, ok 1.  
 Right primer: considered 1, ok 1.  
 Primer pairs: considered 2, no overlap of required point 1, ok 1

- You can read the result to see how this primer fits your template DNA. This result also includes the products on potentially unintended templates. The fewer the unintended templates the more specific the primer is to your desired gene.

**Graphical view of primer pairs**

**Detailed primer reports**

**Primer pair 1**

	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TGGTAGGTCAATCACCCAACC	Plus	21	886	906	59.65	52.38	5.00	5.00
Reverse primer	GCTGCTGTAGTGCCTGATGG	Minus	20	969	950	61.09	60.00	3.00	1.00
Product length	84								

**Products on intended targets**  
 >NM\_166789.2 Drosophila melanogaster eyeless (ey), transcript variant B, mRNA

product length = 84  
 Forward primer 1 TGGTAGGTCAATCACCCAACC 21  
 Template 886 ..... 906

Reverse primer 1 GCTGCTGTAGTGCCTGATGG 20  
 Template 969 ..... 950

**Products on potentially unintended templates**  
 >NM\_001014693.2 Drosophila melanogaster eyeless (ey), transcript variant D, mRNA

product length = 84  
 Forward primer 1 TGGTAGGTCAATCACCCAACC 21  
 Template 1494 ..... 1514

Reverse primer 1 GCTGCTGTAGTGCCTGATGG 20  
 Template 1577 ..... 1558

>NM\_001014694.2 Drosophila melanogaster eyeless (ey), transcript variant C, mRNA

- Save the screenshot of the Graphical view of primer pairs and copy all the detailed primer reports (all the products on intended targets or potentially unintended templates) to the report. Remember to evaluate this primer and give a suggestion. **Be aware to indicate which primer is evaluated and where it comes from.**



## Lab Exercise III - FlyPrimerBank Primer validation.

### Material

Laptop

FlyPrimerBank(<https://www.flyrnai.org/flyprimerbank>)

NCBI Primer-Blast tool (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi>)

### Methods

1. Download a pair of primers from FlyPrimerBank. Students can type in the 'Enter gene identifiers/Primer IDs' box anyone of **vg** , **BH1**, **e**, **e**

## Search FlyPrimerBank for PCR Primers in *D. melanogaster*

[Details](#)

[Documentation / Demo Video](#)

Enter gene identifiers/primer IDs ([help](#)):

ey

Or upload from a file of gene IDs:

2. Click on the search button. Suggested primers will be displayed in seconds. Pick the 1st pair. Save the primer text to your report.

Search Term	Primer Pair	Forward Primer	Fwd Pri Temp	Reverse Primer	Rev Pri Temp	PCR Product	FBgn	Symbol	Isoforms Targeted	Region	Exon Spanning?
ey	PP21231	TGGTAGGTCAATCACCAACC	61.4°C	GCTGCTGTAGTGCCTGATGG	62.9°C	84	FBgn0005558	ey	4/4	CDS	No
ey	PP33244	TCAGGCACTACAGCAGCATC	61.9°C	TTTCGCTAAGAGAGGTGGGGT	62.6°C	83	FBgn0005558	ey	4/4	CDS	No
ey	PP8355	CGGCTGTTAAATACTCAGCACG	61.6°C	TGGGTGATTGACCTACCAAGG	61.1°C	82	FBgn0005558	ey	4/4	CDS	No

3. Validate the FlyPrimerBank primer by using NCBI Primer-Blast tool, aka. pick primers. Leave the PCR template blank, copy-paste the primers from the FlyPrimerBank result and choose the organism (*Drosophila melanogaster* - *taxid:7227*). Submit and wait for the result. This is also a stimulation of the PCR experiment using a certain primer.



Primers for target on one template      Primers common for a group of sequences

Retrieve recent results    Publication    Tips for finding specific primers

### PCR Template

Enter accession, gi, or FASTA sequence (A refseq record is preferred) [?](#) [Clear](#)

Range [?](#) [Clear](#)

Forward primer  From  To

Reverse primer

Or, upload FASTA file    [选择文件](#)    未选择任何文件

### Primer Parameters

Use my own forward primer (5'→3' on plus strand)  [?](#) [Clear](#)

Use my own reverse primer (5'→3' on minus strand)  [?](#) [Clear](#)

PCR product size    Min  Max

# of primers to return   

Primer melting temperatures (T<sub>m</sub>)    Min  Opt  Max  Max T<sub>m</sub> difference  [?](#)

### Exon/intron selection

A refseq mRNA sequence as PCR template input is required for options in the section [?](#)

Exon junction span     [?](#)

Exon junction match    Min 5' match  Min 3' match  Max 3' match

Minimal and maximal number of bases that must anneal to exons at the 5' or 3' side of the junction [?](#)

Intron inclusion    ☐ Primer pair must be separated by at least one intron on the corresponding genomic DNA [?](#)

Intron length range    Min  Max  [?](#)

**Note: Parameter values that differ from the default are highlighted in yellow**

### Primer Pair Specificity Checking Parameters

Specificity check    ☒ Enable search for primer pairs specific to the intended PCR template [?](#)

Search mode     [?](#)

Database     [?](#)

Exclusion    ☐ Exclude predicted Refseq transcripts (accession with XM, XR prefix) ☐ Exclude uncultured/environmental sample sequences [?](#)

Organism     [Add organism](#)

Enter an organism name (or organism group name such as enterobacteriaceae, rodents), taxonomy id or select from the suggestion list as you type. [?](#)

- Save the screenshot of the Graphical view of primer pairs and copy all the detailed primer reports (all the products on intended targets or potentially unintended templates) to the report. Remember to evaluate this primer and give a suggestion. **Be aware to indicate which primer is evaluated and where it comes from.**

### Assignment 7

- Create a word file as your lab 7 report.
- Lab Exercise I : Design primers for **Vestigial, Bar, and Ebony** genes. If the gene has multiple isoforms, design a primer to fit all of them. Follow the Lab Exercise I method and save each result to the report.
- Lab Exercise II: Do primer validation for each pair of primers given by your instructor.
- Lab Exercise III: Do primer validation for one pair of primers chosen by yourself.

**Lab 8 Scientific Methods & Biotechnology and Society (Lecture Only)**  
**Research Project**  
Na Xu

**Objective**

- To be familiar with the scientific method, hypothesis, controls, constants, independent variable, and dependent variable
- To practice experimental design

**Introduction**

Scientific methods are the technique used in the construction and testing of a scientific hypothesis (an educated guess). The scientific method is critical to the development of scientific theories, which explain empirical (experiential) laws in a scientifically rational manner. In a typical application of the scientific method, a researcher develops a hypothesis, tests it through various means, and then modifies the hypothesis on the basis of the outcome of the tests and experiments. The modified hypothesis is then retested, further modified, and tested again, until it becomes consistent with observed phenomena and testing outcomes.

Control groups are the standard to which comparisons are made in an experiment. Many experiments are designed to include the positive control and/or negative groups and one or more experimental groups. The **positive control** contains the variable for which you are testing. It produces a positive reaction and demonstrates the test's ability to detect what you expect. A positive reaction to a positive control demonstrates that your test reacts correctly. A **negative control** does not contain the variable for which you are searching. It contains only the solvent (often distilled water with no solute) and does not react in the test. A negative control demonstrates what a negative result looks like.

Constant is a quantity that does not change. Although you can measure a constant, you cannot alter the constant during an experiment. Contrast this with an experimental independent variable or dependent variable, which is the part of an experiment that is changed.

An Independent variable is a variable that is changed or manipulated in a scientific experiment.

A Dependent variable is the effects on the dependent variable. Its value depends on changes in the independent variable.

A hypothesis is a precise, testable statement of what the researcher(s) predict will be the study's outcome. It is stated at the start of the study. The hypothesis usually involves proposing a possible relationship between the independent variable (what the researcher changes) and the dependent variable (what the research measures). A fundamental requirement of a hypothesis is that it can be tested, and can then be supported or rejected. There are two types of hypotheses: 1) a **null hypothesis** states that there is no relationship between the two variables being studied (one variable does not affect the other). There will be no changes in the

dependent variable due to the manipulation of the independent variable; 2) the **alternative hypothesis** states that there is a relationship between the two variables being studied (the independent variable causes an effect on the dependent variable). The alternative hypothesis states that there is a relationship between the two variables being studied (one variable has an effect on the other). The alternative hypothesis can further be divided into a non-directional or directional hypothesis. A **non-directional (two-tailed) hypothesis** predicts that the independent variable will have an effect on the dependent variable, but the direction of the effect is not specified. It just states that there will be a difference. A **directional (one-tailed) hypothesis** predicts the nature of the effect of the independent variable on the dependent variable. It predicts in which direction the change will take place, i.e. greater, smaller, less, more.

Core elements of good experimental design: 1. **Comparison/control/baseline:** In an experiment where you're measuring the *dependent variable*, the effect of treatment with the *independent variable*, it is important to assess that effect without having the *independent variable*. It is a good practice to always include a *negative control* and *positive control* in your experimental design so that your data have a baseline to compare to. The controls also prove that your experimental materials are used properly and procedures are conducted correctly. 2. **Blocking/controlling for confounds:** In an experiment with possible external sources of variability other than your *independent variable*. In statistics, a **confounder** is a variable that influences both the *dependent variable* and *independent variable*, causing a possible association. Controlling for *confounds* is best achieved through blocked design, where we divide subjects into groups corresponding to levels of a confounding factor, and repeat the experiment for each group. By accounting for the effect of confounding variables, we can avoid being misled by our data. For example, an educational intervention program may have different effects on students of a different gender. Gender in this case is a *confounder* that could influence our results and we should divide students by gender when conducting our study. 3. **Replication:** Any good experiment should be reproducible, and in particular, replication should yield similar results. Increasing sample numbers and repeating experiments at least three times will improve the reproducibility and accuracy of results.

Placebo effect: When you apply a treatment to a subject, that treatment may be theoretically ineffective, but can still produce a significant effect in reality simply due to the existence of the treatment. For example, a fake placebo surgery can actually do as well as common knee surgery. And giving people nonalcoholic drinks but telling them that the drinks are alcoholic can result in a decline in memory powers. Because of the placebo effect, the development of medical treatments demands a stronger standard of outperforming a placebo rather than outperforming not giving any treatment at all since a placebo alone could already result in a startling improvement in a test subject's condition, often largely due to psychological factors. To examine the benefit of a placebo, an experiment could have a control group that receives no treatment, a placebo group that receives a sham treatment, and a treatment group that receives the actual treatment under study. By doing this, the experimenter can measure both the effect of the placebo over the baseline and the effect of the treatment over the placebo.

Hawthorne effect: The Hawthorne effect is a type of reactivity in which individuals modify an aspect of their behavior in response to their awareness of being observed. In a famous experiment in the early 20th century, a factory called the Hawthorne Works wanted to measure the effect of lighting on productivity. An experimental group had their light bulbs changed, and the experimenters wanted to measure the effect on productivity. A control group saw workers change their bulbs, but the new bulbs were identical to the old ones. However, both groups improved after the “new” bulbs were put in: the control group improved purely due to their perception of the effect.

## **Exercise I - Experimental Design Practice**

Activity One: Identifying and creating scientifically testable hypotheses; practicing good experimental design

1. Students will evaluate the following hypothesis and formulate experimental designs to test those hypotheses. While designing experiments, students should identify and include proper controls, independent variables, dependent variables, constants, and confounders. Students should also consider the potential placebo effect and Hawthorne effect.
  - 1). Classic music helps plants to grow more.
  - 2). Doing exercise helps people to be healthier.
  - 3). Men are stronger than women.
  - 4). DNA was brought to the planet by aliens.
2. Students will formulate a hypothesis based on the following case study and design experiments to test the hypothesis. While designing experiments, students should identify and include proper controls, independent variables, dependent variables, constants, and confounders. Students should also consider the potential placebo effect and Hawthorne effect.

Case Study: During the 2004 Tsunami disaster in Asia, there were reports that trained elephants had moved away from the coastline before the tsunami struck. Scientists wondered if elephants may have sensed the oncoming wave and earthquake that created it. Elephant trunks have two kinds of nerve endings that are sensitive to pressure differences, but little is known about nerve ending distribution in the elephant’s skin or feet. An elephant has recently died at a nearby zoo and scientists have decided to dissect one of its feet to reveal any special nerve endings.

Results indicate a high density of pressure-sensitive nerve endings in the front of the foot and along the edges, which supports the idea that elephants are sensing seismic signals produced by earthquakes.

### Activity Two: Evaluating Experimental Design

Case study: The Food and Drug Administration asked two groups of scientists to design an experiment to test a new drug that is supposed to help people lose weight. Read the designed experiments by these two groups of researchers and answer the following questions.

Group 1: 50 women and 50 men with BMI (body mass index) between 30-33% were chosen as the participants for this experiment. All participants were the same in terms of health, age and ethnicity. All the participants received instructions on their diets, exercise, and sleep time before starting the experiment. During the 12-week duration of the experiment, half the participants (25 women and 25 men) received identical pills without medication (called placebos). None of the participants in either group knew which pill they were receiving. The scientists who analyzed the data from this experiment did not know which participant received either the real medication or placebo.

Group 2: 50 women and 50 men with BMI (body mass index) between 30-33% were chosen as the participants for this experiment. The participants were in the same range of age. All participants received instructions on their diets, exercise, and sleep time before starting the experiment. During the 12-week duration of the experiment, half the participants (25 women and 25 men) received medication in pill form twice per day every 12 hours. The other half of the participants did not receive any medication.

1. Which experiment design is more reliable between the two groups? Give three reasons
2. Which was the control group for each of these experimental designs? Which group designed the better control group? Why?
3. What were the independent variable and the dependent variable?
4. What confounders were included in this experimental design?
5. Why it was important that neither the participants nor the scientists knew which participant actually took the weight loss medication?

## Exercise II - Formulate Hypothesis and Design Experiment Using *Drosophila*

Background information: Diet-mediated diseases represent an increasing challenge in Western society. Particular attention has focused on carbohydrate consumption, which has increased as significantly in the past three decades. Dietary sugars have in turn been linked to a variety of metabolism-related problems including obesity, insulin resistance, metabolic syndrome, and type 2 diabetes mellitus (T2DM), as well as cardiovascular disease. *Drosophila melanogaster* provides a useful complement to mammalian models. While flies show important differences from their mammalian counterparts, they also show marked similarities. For example, the fly genome encodes seven insulin-like peptide genes (Dilp 1–7) that activate classical insulin pathway signaling. Ablation of insulin-producing cells phenocopied several aspects of type 1 diabetes. Further, *Drosophila* is shown to be susceptible to diet-mediated metabolic and cardiac organ dysfunctions that are reminiscent of those reported in mammals. Therefore, *Drosophila* serves as a proper model system to study the impact of dietary sugar on physiology and behavior.

The objective of the study: Students will work in groups to identify the impact of increased dietary sugar on *Drosophila*'s weight, life span, and reproduction.

Group activity: Students are divided into 3-4 students per group, discuss, formulate a testable hypothesis, and draft a research plan, including proper controls, independent variables, dependent variables, constants, and confounders. Then students will start conducting the research, collect data and report findings during the next few labs.

### Assignment 8:

Students will formulate a testable hypothesis, and draft a research plan, including proper controls, independent variables, dependent variables, constants, and confounders.

## Biotechnology and Society Module I: Genetically Modified Organisms (GMO)

For thousands of years, humans have been using traditional modification methods like **selective breeding and cross-breeding** to breed plants and animals with more desirable traits. For example, early farmers developed cross-breeding methods to grow corn with a range of colors, sizes, and uses. Today's strawberries are a cross between a strawberry species native to North America and a strawberry species native to South America.

Changing plants and animals through traditional breeding can take a long time, and it is difficult to make very specific changes. After scientists developed genetic engineering in the 1970s, they were able to make similar changes in a more specific way and in a shorter amount of time. You can read more history about major events about “Genetic Modification in Modern Agriculture” through this link: <https://www.fda.gov/media/135276/download>.

**Genetically modified (GM) foods** are foods derived from organisms whose genetic materials are modified in a way that does not occur naturally. Transferring genes into crop plants confers resistance to herbicides, insect pests, and plant diseases. Gene transfer can also be used to increase the nutritional value of foods. The technology used is called “**recombinant DNA technology**” or “**genetic engineering**”. Currently, available GM foods stem mostly from plants, but in the future foods derived from GM microorganisms or GM animals are likely to be introduced on the market. GM foods can also allow for reductions in food prices through improved yields and reliability. All GM foods should be assessed before being allowed on the market. FAO/WHO Codex guidelines exist for risk analysis of GM food. All genetically modified foods currently available on the international market have passed safety assessments. When discussing and developing GM foods, three main safety issues are considered. The first is allergenicity, meaning the ability of the edited genes or food product to cause an allergic reaction. No allergic effects have been found relative to GM foods currently on the market. The second is the transfer of genes from GM foods to the human digestive tract. The third is outcrossing, meaning the transfer of genes from GM organisms to other species. Several countries have developed preventative strategies, including a clear separation of GM and non-GM food crops.

Nowadays you can easily find many GM food in the grocery store: soybeans, cotton, canola, and sugar beets in the market have been genetically modified to be naturally resistant to an herbicide. This increased resistance to the herbicide allows farmers to use more herbicides to kill weeds. Most cotton has been genetically modified to increase yield and resistance to disease. In 2018, GMO (genetically modified organisms) soybeans made up 94% of all soybeans planted, GMO cotton made up 94% of all cotton planted, and 92% of corn planted was GMO corn. In 2013, GMO canola made up 95% of canola planted while GMO sugar beets made up 99.9% of all sugar beets harvested. Most GMO plants are used to make ingredients that are then used in other food products, such as cornstarch made from GMO corn or sugar made from GMO sugar beets. GMO soy is also used as an ingredient (lecithin, emulsifiers, and proteins) in processed foods.

Corn is the most commonly grown crop in the United States, and most of it is GMO. Most GMO corn is created to resist insect pests or tolerate herbicides. *Bacillus thuringiensis* (Bt) corn is GMO corn that produces proteins that are toxic to certain insect pests but not to humans, pets, livestock, or other animals. These are the same types of proteins that organic farmers use to control insect pests, and they do not harm other, beneficial insects such as ladybugs. GMO Bt corn reduces the need for spraying insecticides while still preventing insect damage. While a lot of GMO corn goes into processed foods and drinks, most of it is used to feed livestock, like cows, and poultry, like chickens.



By the 1990s, ringspot virus disease had nearly wiped out Hawaii's papaya crop. GMO papaya, named the Rainbow papaya, was created to resist the ringspot virus. This GMO saved papaya farming on the Hawaiian Islands.

A few varieties of GMO apples were developed to resist browning after being cut. This helps cut down on food waste, as many consumers think brown apples are spoiled.

More than 95% of animals used for meat and dairy in the United States eat GMO crops. The DNA in the GMO food does not transfer to the animal that eats it. This means that animals that eat GMO food do not turn into GMOs. Similarly, the DNA from GMO animal food does not make it into the meat, eggs, or milk from the animal. Research shows that foods like eggs, dairy products, and meat that come from animals that eat GMO food are equal in nutritional value, safety, and quality to foods made from animals that eat only non-GMO food. **GMO animals** will soon be available for food supply. FDA has approved an application that allows the marketing of the AquAdvantage Salmon, an Atlantic salmon that has been genetically modified to reach an important growth point faster. FDA determined that AquAdvantage Salmon is as safe to eat and as nutritious as non-GMO Atlantic salmon.

**Concerns about GMOs may include:**

- Are foods containing new proteins safe to eat?
- Is it Safe to eat food carrying part of a viral gene that switches on transgenes?
- Will pesticide-resistant insects develop?
- Will disease-causing bacteria acquire antibiotic resistance genes used as markers?

*\*Students will use the material in the "Biotechnology and Society" modules to complete Biotechniques Reflection Assignment" for this course.*



## Lab 9 Introduction to Scientific Literature

### Research Project (Continued)

#### Objective

- To be familiar with the scientific literature, resources, and citations
- Set up the experiment for the group research project

#### Introduction

Sharing the results of scientific experiments is one of the most important aspects of science. Our current knowledge of science and the world around us is dependent upon current scientists sharing their results. Scientists share their discoveries with the public through scientific publications in journals and books. Other scientists can then find this information to help them design their own experiments, put their own research in context, and to add to our knowledge of the world. One of the most common ways for scientists to formally share their research is to publish a **“journal article.”** The journal article is a part of the formal written record of the scientific process and results. A credible journal is usually peer-reviewed. **Peer-reviewed publications** are written by scholars for scholars and are reliable sources of information on a topic or discipline. These journals can be found either in the library's online databases or in the library's local holdings. **Peer review** is a process where an article is verified by a group of scholars before it is published. When an author submits an article to a peer-reviewed journal, the editor passes out the article to a group of scholars in the related field (the author's peers). They review the article, making sure that its sources are reliable, and that the information it presents is consistent with the research. The reviewers determine whether the article will be accepted for publication, need more revisions, or be rejected. The peer-review process makes sure that only quality research is published. There are two types of scientific articles that can be published: **review articles** and **original research articles**. If the authors are discussing research and experiments that they carried out, and giving an outline of the experiment, it will be an **original research article**. In **review articles**, authors are discussing experiments done by other scholars. Review articles usually summarize the findings of many original research articles about a focused topic, whereas original research articles show the original research methods, materials, and data.

To search for credible scientific publications, we usually recommend you to use the **“OneSearch”** at LaGuardia’s online library (<https://library.laguardia.edu/>) or **PubMed**, a free search engine accessing primarily the MEDLINE database of references and abstracts on life sciences and biomedical topics, developed by the United States National Library of Medicine at the National Institutes of Health. When you are searching for information on a research topic using popular search engines, you may run across some unreliable resources. Shorter news articles may appear in some scientific and popular publications reporting. These news articles are not peer-reviewed and are normally written by science journalists, not researchers. The news articles may be easier to read, but since they are not the original research, and therefore, a news article may not be the best source for your paper or project. However, news articles can lead you to a piece of original research and can help you easily stay informed about recent

research developments. If you conduct your searches online, via Google, Yahoo, or another popular search engine, you may find journal articles, but you may also come across other scientific information that can take many forms. Wikis, blogs, and personal websites can often contain a lot of scientific information, but these resources are generally very far removed from the original research where the ideas were first developed. Each of these sources needs to be evaluated very carefully to determine if the information is credible, and these sources won't be suitable for a research paper.

A **scientific literature review** is a critical account of what has been published on a topic by accredited researchers. We often need to write a scientific literature review for assignments, an introduction to an essay, a report, a thesis, and a part of research or grant proposals. Writing a literature review will improve your topic knowledge, provide new insight on your topic to readers, and help you develop literature searching abilities, critical analysis skills, and communication and writing skills. Writing a scientific literature review needs to follow the styles of scientific writing. The writing needs to be precise, clear, and objective. Simple language should be used.

## **Biotechnology and Society Module II: Genetic Testing and Screening**

**Genetic testing** is used to determine if someone has a genotype or is a carrier for a genetic disorder. **Newborn screening** is a set of laboratory tests performed on newborn babies to detect a set of known genetic diseases. Typically, this testing is performed on a blood sample obtained from a heel prick when the baby is two or three days old. In the United States, newborn screening is mandatory for a defined set of genetic diseases, although the exact set differs from state to state. Mandatory newborn screening programs are a classic public health success story. These programs save lives and prevent disability in thousands of infants every year by identifying certain conditions promptly after birth. Early detection of those conditions allows different kinds of treatments and interventions to be provided very early in a newborn's life when it can really make a difference in health outcomes and even help to prevent severe, irreversible disability. New York state offers a newborn screening panel for 56 conditions (<https://newbornscreening.hrsa.gov/your-state/new-york>).

**Carrier testing** searches for heterozygotes that may be at risk of producing a defective child. When it is done before or during pregnancy, it allows you to find out your chances of having a child with a genetic disorder. Carrier screening is available for a limited number of diseases, including cystic fibrosis, fragile X syndrome, sickle cell disease, hemoglobinopathies, spinal muscular atrophy, and Tay-Sachs disease. Some of these disorders occur more often in certain races or ethnic groups. For example, sickle cell disease often affects people of African descent. Tay-Sachs disease typically affects people of Eastern or Central European Jewish, French Canadian, and Cajun descent.

**Prenatal testing** tests a fetus for genetic disorders such as cystic fibrosis or chromosome abnormalities such as Down Syndrome. Prenatal testing is offered in all pregnancies. A cell-free DNA (cfDNA) screening is offered as early as week 10 of pregnancy through a blood test,

evaluating the risk of Down syndrome (trisomy 21), trisomy 18 and trisomy 13, and fetal sex. The first-trimester screening includes a blood test and ultrasound to evaluate the risks of Down syndrome (trisomy 21) and trisomy 18. A quad screening is done through a blood test at weeks 15-22 of pregnancy to evaluate the risk of Down syndrome (trisomy 21), trisomy 18, and open neural tube defects, such as spina bifida. In addition, fetal ultrasounds also play an important role in prenatal screening. Prenatal testing is associated with some risks such as infection, hemorrhage, fetal injury, and spontaneous abortion.

**Preimplantation genetic testing (PGT)** is a screening test that can be performed on embryos created via in vitro fertilization (IVF) to genetically analyze the embryos prior to transfer and implantation. IVF is a requirement in order to obtain embryos for PGT. After eggs are retrieved and fertilized in the lab, the embryologist will perform assisted hatching on the embryos to help obtain cells for testing. When the embryo reaches the blastocyst stage, a few of the cells on the outside of the embryo are biopsied and sent for genetic analysis.

**Presymptomatic gene testing** is genetic testing performed on an individual who does not show symptoms of the disorder, but who is at risk of developing the disorder. Presymptomatic DNA testing for autosomal dominant hereditary breast/ovarian cancer (HBOC) became an option after the identification of the *BRCA1* and *BRCA2* genes in 1994-1995. Healthy female mutation carriers have a high lifetime risk for breast cancer (56-87%) or ovarian cancer (10-60%) and may opt for intensive breast and ovary surveillance or prophylactic surgery (mastectomy/oophorectomy).

*\*Students will use the material in the "Biotechnology and Society" modules to complete Biotechniques Reflection Assignment" for this course.*

### **Exercise I- Preparing Fly Food with High Dietary Sugar**

The food protocol is adapted from the protocol used in lab 2. Standard fly food was prepared with an additional 1%, 4%, 10%, and 15% sucrose. The control is the standard fly food without any additional sucrose added. Students are divided into 5 groups and each group is responsible to make one type of fly food including the control.

#### Materials

Hotplate  
Nutri-Fly food  
Distilled water  
Balancer  
1 L and 100 mL beakers  
Spatula  
Microwave  
Thermometer

Heat resistant gloves  
Tegosept  
Ethanol  
Phosphoric Acid  
Fly food bottles, vials, and trays  
Cotton and foam plugs  
Funnel  
Live yeast dilution  
Sucrose

### Methods

1. Preheat the hotplate at setting 3.
2. Weigh 35.6 grams of Nutri-Fly food, and add no sucrose or add a specific amount of sucrose:
  - 1% sucrose: add 2 g sucrose
  - 4% sucrose: add 8 g sucrose
  - 10% sucrose: add 20 g sucrose
  - 15% sucrose: add 30 g sucrose
3. While the plate is heating, place a 500 mL beaker on the plate, add 200 mL of distilled water, and slowly pour in the Nutri-Fly food with added sugar. Use a spatula constantly to break up any food bubbles or lumps.
4. Microwave for 2 min and then stir with a spatula. Repeat the second time for a total of 4 min.
5. Microwave for an additional 30-45 seconds without letting the food boil over. (*Watch the meniscus carefully. When it starts to rise, stop the microwave right away*)
6. Place the beaker on a hotplate, cover with foil, and stir. The food should be at the boiling stage, making some bubbles steadily, but not boiling over the container. During this time, insert a thermometer through the foil cover such that the bulb is just below the surface but clear of the stir bar. Set up a timer for 10 min.
7. During the 10 min waiting, prepare a 10% Tegosept (dissolve 1.5 g Tegosept in 15 mL ethanol). Tegosept is an anti-fungal drug to prevent fungi infection in fly cultures. This is a Tegosept stock to be shared by 6 groups of students
8. After 10 min, put the food on the bench. Keep stirring until the temperature drops to 75 °C.
9. Remove the thermometer, each group add 2 mL of 10% Tegosept from the prepared stock at step 7 and 68 uL of Phosphoric Acid and mix thoroughly by hand using a spatula. The fly food is ready and needs to be poured into containers quickly before they solidify.

Phosphoric Acid is used to acidify food and has fungicidal effects. Phosphoric Acid is a strong acid and corrosive. Gloves are needed when handling Phosphoric Acid.

10. Pour about 5 ml per vial. Wipe excess food off the sides of the vial/bottle with tissue paper.
11. Stack the trays onto each other or cover the top tray with paper towels to keep flies from the food before you can plug.
12. Leave at room temperature overnight to dry. Label the percentage of sugar added.
13. The next day, plug in cotton and place food in plastic bags. Store at room temperature for short-term use and in the refrigerator for long-term use (this will be completed by the lab technician or instructor)

### **Assignment 9**

Students will conduct and write a scientific literature review about their proposed hypothesis. The minimal word requirement is 500 words.

### **Oral Presentation Assignment**

Students will also make a 5-10 min **oral presentation video** for the scientific literature review and their research goal. The oral presentation video counts for 10% of the final grade

## Lab 10 Methodology, Data Analysis, and Results of Research Papers, Biotechnology and Society & Research Project (Continued)

Na Xu

### Objective

- To be familiar with the Methodology, Data Analysis, and Results sessions of a research paper
- Continue experiment for the group research project

### Introduction

*The methods section* of a research paper describes how experiments are conducted, provides the information by which a study's validity is judged. The method section answers three main questions: 1) How was the data collected or generated? 2) What materials were used? and 3) How were data analyzed? The writing should be direct and precise and written in the past tense.

#### **An effectively written methodology section should:**

- **Introduce the overall methodological approach for investigating your research problem.** You need to include all the materials used and list all the steps of the experiments.
- **Indicate how the approach fits the overall research design.** Your methods should have a clear connection with your research problem. In other words, make sure that your methods will actually address the problem. You should provide background and rationale for methodologies that are unfamiliar to your readers. You should also provide a rationale for subject selection and sampling procedure. For instance, if you propose to conduct interviews, how do you intend to select the sample population? If you are analyzing texts, which texts have you chosen, and why? If you are using statistics, why is this set of statistics being used? If other data sources exist, explain why the data you chose is most appropriate. You need to state clearly the controls and variables in your experiment and explain the rationale for using this particular controls and variables.
- **Describe the specific methods of data collection you used,** such as experiments, surveys, interviews, questionnaires, observation, and archival research. If you are analyzing existing data, such as a data set or archival documents, describe how it was originally created or gathered and by whom.
- **Explain how you intend to analyze your results.** You should include any statistical analysis, software, and tools used to analyze data, as well as the rationale for using those tools.
- **Address problems and limitations.** It is almost a given that you will encounter problems when collecting or generating your data. Do not ignore these problems or pretend they did not occur. Often, documenting how you overcame obstacles can form an interesting part of the methodology. It demonstrates to the reader that you can provide a cogent rationale for the decisions you made to minimize the impact of any problems that arose. Nearly all methodology has its limitations. Are there any practical limitations that could affect your data collection? How will you attempt to control for potential confounding

variables and errors? You should state openly the problems that may arise due to the methodology.

Data analysis is a technique that typically involves multiple activities such as collecting, evaluating, and organizing the data. Data analysis is usually shown in the “**Results**” session of your paper. The **Results** section of the research paper is where you report the findings of your study based upon the information gathered as a result of your experiments. The results section should simply state the findings, without bias or interpretation, and be arranged in a logical sequence. The results section should always be written in the past tense. A section describing results or findings is particularly necessary if your paper includes data generated from your own research. **In general, the content of your results section should include the following elements:**

- An introductory context for understanding the results by restating the research problem that underpins the purpose of your study.
- Inclusion of **non-textual elements**, such as figures, charts, photos, maps, tables, etc. to further illustrate the findings, if appropriate.
- Inclusion of **figure legends** explaining the non-textual elements presented.
- In the text, a systematic description of your results, highlighting for the reader observations that are most relevant to the topic under investigation.
- Use the past tense when referring to your results.

Using Non-textual elements: You can place figures, tables, charts, etc. within the text of the result or in the end of the report as a separate session called “Figures”. Below are several rules while using non-textual elements in your report.

- In the text, refer to each non-textual element in numbered order (e.g., Table 1, Table 2; Chart 1, Chart 2; Map 1, Map 2; Figure 1, Figure 2).
- Each non-textual element must be titled, numbered consecutively, and complete with a figure legend explaining the results
- In proofreading your results section, be sure that each non-textual element is sufficiently complete so that it could stand on its own, separate from the text.
- Incorporate non-textual elements well into the textual writings.

Problems to avoid: When writing the results section, avoid doing the following:

- **Ignoring negative results.** If some of your results fail to support your hypothesis, do not ignore them. Document them, then state in your discussion section why you believe a negative result emerged from your study. Note that negative results, and how you handle them, often provide you with the opportunity to write a more engaging discussion section, therefore, don't be afraid to highlight them.
- **Including raw data or intermediate calculations.** All the data in the Result session should be analyzed, and organized. Ask your professor or supervisor if you need to include any raw data generated by your studies, such as transcripts from interviews or data files. If raw data is to be included, place it in an **Appendix** or **Supplementary Material** session that is referred to in the text.

- **Be vague, not confident, or non-specific.** Do not use phrases that are vague or non-specific, such as, "appeared to be greater or lesser than..." or "demonstrates promising trends that..."
- **Presenting the same data or repeating the same information.** If you feel the need to highlight something, you will have a chance to do that in the discussion section.
- **Discuss your results.** Save all this for the **"Discussion"** section of your paper.

### Exercise I - Research Project Continued

Students will work in groups to continue their research in investigating the impact of increased dietary sugar on *Drosophila*'s weight, life span, and reproduction.

#### Materials

Prepared fly food with 0, 1%, 4%, 10%, or 15% sucrose

Wild-type fly culture that is ready for newborn male and female collections

Stereo microscopes

Forceps

Fly nap

#### Experimental Procedures:

1. Students select 5 newborn males and 5 virgin female wild-type flies and transfer the cross into new food vials containing 0, 1%, 4%, 10%, or 15% sucrose (5 males and 5 females in each vial). The food vial that does not contain any additional sucrose serves as the negative control.
2. After setting up the crosses. Students label the dates. All crosses will be incubated at room temperature for two weeks to collect results

### Assignment 10

Students will draft the Methodology section of their research project.

### Biotechnology and Society Module III: Genome Editing Technology in the Targeted Therapy of Human Diseases

**Genome Editing**, also called **Gene Editing**, is an area of research seeking to modify genes of living organisms to improve our understanding of gene function and develop ways to use it to treat diseases. Genome editing can be used to correct, introduce or delete almost any DNA sequence in many different types of cells and organisms. While techniques to modify DNA have existed for several decades, new methods have made genome editing faster, cheaper, and more efficient. Genome editing builds on an earlier discovery that a broken section of DNA in a gene triggers a cell's repair mechanism to stitch together the break. Genome editing allows researchers to mimic this natural process of DNA repair.



**Advanced genome editing methods** include zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), meganucleases, as well as the newly developed method, clustered regularly interspaced short palindromic repeats, and CRISPR-associated protein 9 (CRISPR/Cas9).

**CRISPR/Cas9** is the most widely used genome editor and is a powerful tool for understanding gene function. The CRISPR-Cas9 system is faster, cheaper, more accurate, and more efficient than other genome editing methods. CRISPR-Cas9 was adapted from a naturally occurring genome editing system that bacteria use as an immune defense. When infected with viruses, bacteria capture small pieces of the viruses' DNA and insert them into their own DNA in a particular pattern to create segments known as **CRISPR arrays**. The CRISPR arrays allow the bacteria to "remember" the viruses (or closely related ones). If the viruses attack again, the bacteria produce **RNA segments** from the CRISPR arrays that recognize and attach to specific regions of the viruses' DNA. The bacteria then use **the Cas9** enzyme to cut the virus DNA and disable the virus. Scientists adapted the bacterial immune defense system to edit DNA. Scientists create a **guide RNA**, a small piece of RNA with a short "guide" sequence that attaches (binds) to a specific target sequence in a cell's DNA, much like the RNA segments bacteria produce from the CRISPR array. This guide RNA also attaches to the Cas9 enzyme. When introduced into cells, the guide RNA recognizes the targeted DNA sequence, and the Cas9 enzyme cuts the DNA at the targeted location, mirroring the process in bacteria. Once the DNA is cut, researchers use the cell's own **DNA repair machinery** to add or delete pieces of genetic material, or to make changes to the DNA by replacing an existing segment with a customized DNA sequence (Gupta et al., 2014, Hsu et al., 2014)). The CRISPR-Cas9 system is being explored in research and clinical trials for a wide variety of diseases, including single-gene disorders such as cystic fibrosis, hemophilia, and sickle cell disease. It also holds promise for the treatment and prevention of more complex diseases, such as cancer, heart disease, mental illness, and human immunodeficiency virus (HIV) infection.

**Ethical concerns** arise when genome editing, using technologies such as CRISPR-Cas9, is used to alter human genomes. NIH (National Institute of Health) supports human genome editing approaches in somatic cells and supported *the Somatic Cell Genome Editing Program*. Somatic cells are not involved in human reproduction, and gene editing is not passed from one generation to the next. However, changes made to genes in egg or sperm cells or to the genes of an embryo could be passed to future generations. Germline cell and embryo genome editing bring up a number of ethical challenges, including whether it would be permissible to use this technology to enhance normal human traits (such as height or intelligence). Based on concerns about ethics and safety, germline cell and embryo genome editing are currently illegal in the United States and many other countries.

CRISPR and other gene editing methods, especially ZFNs, are speeding gene therapy approaches to treat many human conditions. In 2014, the first clinical application of genome editing involved the use of ZFNs to make human cells resistant to HIV-1 by disrupting a gene required for the virus to infect cells. CCR5 is an HIV-1 co-receptor required for virus entry. The

clinical trial reconstituted HIV-1 patients with autologous CD4+ T cells that had been subject to targeted CCR5 mutation using a designer ZFN gene editing system (Tebas et al., 2014).

In 2017, a clinical trial testing ZFNs to correct Hunter syndrome (MPS II) was launched. Hunter syndrome is a lysosome storage disorder. Children with this condition have an abnormal accumulation of complex sugars in their cells, which affects many systems in their bodies. Hunter syndrome is caused by an X-linked recessive gene, the iduronate 2-sulfatase (IDS) gene, which is responsible to make the lysosome enzyme I2S. I2S helps to break down complex sugars called glycosaminoglycans (GAGs). Genetic mutations in the IDS gene result in a deficiency or a complete absence of I2S, which in turn results in an abnormal accumulation of GAGs in the body's cells. Hunter syndrome can cause abnormalities in the skeleton, heart, and respiratory system in children. The clinical trial was the first genome editing approach administered directly to research participants. The ZFN system is used to insert the correct copy of the human IDS (hIDS) coding sequence into the patient's genome and provide long-term expression of IDS.

In 2017, **CAR-T (Chimeric antigen receptor T cells) immunotherapy** was approved by FDA to treat blood cancers. The immune system is highly regulated to avoid attacking healthy cells. T cell PD-1 receptor comes in contact with PD-L1 molecule on another cell and recognizes the cell as a "safe" cell and the T cell leaves the cell alone. Cancer cells express a lot of PD-L1 to prevent T cells attack. For the CAR\_T treatment, scientists harvest T cells from a patient's blood and use CRISPR to mutate the PD-1 receptor gene in T cells so that the T cells gain more capability to recognize, attack, and eliminate cancer cells.

*\*Students will use the material in the "Biotechnology and Society" modules to complete Biotechniques Reflection Assignment" for this course.*

## Lab 11 Protein Analysis Using ELISA

### Objective:

- To explore infectious diseases using a hands-on simulated ELISA assay.
- To understand the basic principles of antibody-mediated immunity.
- To understand the steps used in an ELISA assay and how it is used as a diagnostic tool.
- To understand the disease-causing agent and transmission patterns of certain infectious diseases.

### Introduction

Pathogens: Pathogens are tiny, disease-causing agents including viruses, bacteria, protozoa, molds, and other microorganisms. Pathogens invade the body and multiply, causing sickness or even the death of the infected individual.

The Body's Defense System: The body has three lines of defense to fight infections. The first two modes are nonspecific: the body's physical barriers and the nonspecific immune system. The third defense mechanism is the specific immune system that is tailored to the type of invading pathogen. The external protective covering of the body includes skin, the mucous membranes that line the mouth, nostrils, and other potential gateways. Secretions from the skin and mucous membranes such as sweat, tears, and saliva also prevent pathogens from entering the body. The human body has a series of nonspecific defenses that make up the **innate immune system**. These defenses are not directed against any one pathogen but instead, provide a guard against all infections. If the exterior barriers are breached, nonspecific immune defenses are launched, such as fever. Many pathogens cannot function well at higher body temperatures. An **inflammatory response** begins when a pathogen stimulates an increase in blood flow to the infected area. Blood vessels in that area expand, and white blood cells leak from the vessels to invade the infected tissue. These white blood cells, called *phagocytes* engulf and destroy bacteria. The area often becomes red, swollen, and painful during an inflammatory response. Cells involved in the innate immune system encompass phagocytic cells, epithelial and endothelial cells, natural killer cells, innate lymphoid cells, and platelets. Phagocytic cells consist of granulocytes (i.e., neutrophils, eosinophils, basophils, and mast cells), monocytes/macrophages, and dendritic cells. These cells participate in not only the phagocytosis but also the inflammatory process.

When pathogens are able to bypass innate immune defenses, the **adaptive immune system** is activated. The innate immune system can distinguish between different classes of pathogens and recruit the most effective form of adaptive immune response to eliminate them. Unlike innate immune responses, the adaptive responses are highly specific to the particular pathogen that induced them. They can also provide long-lasting protection. Specific immune responses are triggered by antigens. **Antigens** are usually found on the surface of pathogens and are unique to that particular pathogen.

**Adaptive immune** responses are carried out by white blood cells called lymphocytes. There are two broad classes of such responses—*antibody responses* and *cell-mediated immune responses*,

and they are carried out by different classes of lymphocytes, called B cells and T cells, respectively (Figure 1). B cells are produced and mature in the bone marrow. In **antibody responses**, B cells are activated to secrete antibodies that circulate in the bloodstream and permeate the other body fluids and bind specifically to the foreign antigen (Figure 1). The binding of antibodies inactivates viruses and microbial toxins by blocking their ability to bind to receptors on host cells. Antibody binding also marks invading pathogens for destruction, mainly by making it easier for phagocytic cells of the innate immune system to ingest them.

T cells are also produced in bone marrow, but they mature in the thymus. In cell-mediated immune responses, activated T cells react directly against a foreign antigen that is presented to them on the surface of a host cell (Figure 1). The T cell, for example, might kill a virus-infected host cell that has viral antigens on its surface, thereby eliminating the infected cell before the virus has had a chance to replicate. In other cases, the T cell produces signal molecules that activate macrophages to destroy the invading microbes that they have phagocytosed.

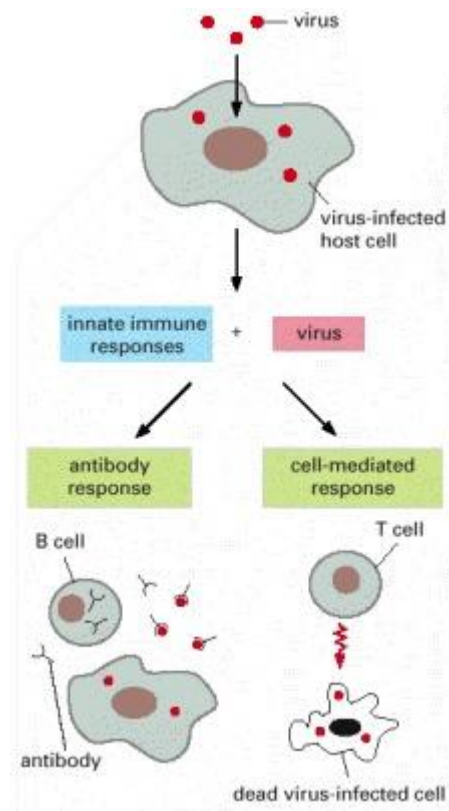


Figure 1: The Two Main Classes of Adaptive Immune Responses. Lymphocytes carry out both classes of responses. Here, the lymphocytes are responding to a viral infection. In one class of response, B cells secrete antibodies that neutralize the virus. In the other, a cell-mediated response, T cells kill the virus-infected cells (Alberts et al., 2002).

ELISA (Enzyme-Linked Immunosorbent Assay) is an immunological assay commonly used to measure antibodies, antigens, proteins, and glycoproteins in biological samples. ELISA assays are based on the principles that antibodies are produced in response to infection and that these antibodies are designed to specifically target particular antigens and bind tightly to them. ELISA is applied widely in diagnosis including diagnosis of HIV infection, pregnancy tests, and measurement of cytokines or soluble receptors in cell supernatant or serum. In a direct ELISA assay, the antigen is immobilized to the surface of the multi-well plate and detected with an antibody specific for the antigen. The antibody is directly conjugated to HRP or other detection molecules. The Direct ELISA Procedure can be summarised into 4 steps: 1) Plate Coating with antigen, 2) Blocking any remaining Protein-binding sites, 3) Antibody Incubation, and 4) Detection.

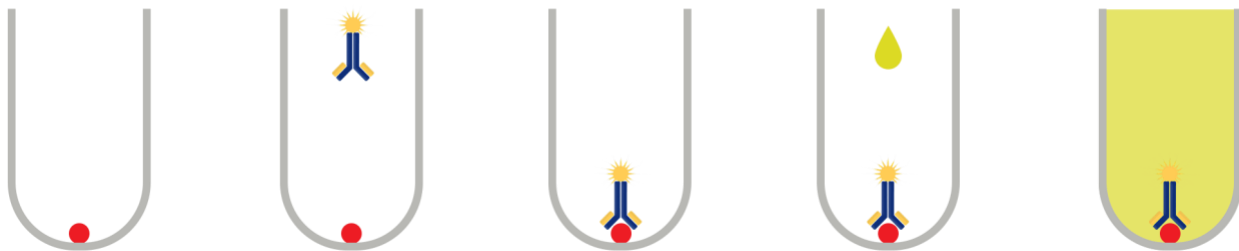


Figure 2: Steps of Direct ELISA Assay (<https://www.abbexa.com/>)

HIV (human immunodeficiency virus) is a virus that attacks the body's immune system, making a person more vulnerable to other infections and diseases. It is spread by contact with certain bodily fluids of a person with HIV, most commonly during unprotected sex, or through sharing injection drug equipment. If HIV is not treated, it can lead to AIDS (acquired immunodeficiency syndrome). There is currently no effective cure. Once people get HIV, they have it for life. But with proper medical care, HIV can be controlled.

Lyme disease is caused by the bacterium *Borrelia burgdorferi* and rarely, *Borrelia mayonii*. It is transmitted to humans through the bite of infected black-legged ticks. Typical symptoms include fever, headache, fatigue, and a characteristic skin rash called erythema migrans.

Avian influenza (bird flu) is caused by infection with avian (bird) influenza (flu) Type A viruses. These viruses naturally spread among wild aquatic birds worldwide and can infect domestic poultry and other bird and animal species. Bird flu viruses do not normally infect humans. However, sporadic human infections with bird flu viruses have occurred. Symptoms begin within two to eight days and can seem like the common flu. Cough, fever, sore throat, muscle aches, headache, and shortness of breath may occur. The disease can carry high mortality in humans. Some antiviral drugs, if taken within two days of symptoms, may help.

West Nile Virus (WNV) is the leading cause of mosquito-borne disease in the continental United States. It is most commonly spread to people by the bite of an infected mosquito. There are no vaccines to prevent or medications to treat WNV in people. Fortunately, most people infected with WNV do not feel sick. About 1 in 5 people who are infected develop a fever and other

symptoms such as headache, body aches, joint pains, vomiting, diarrhea, or rash. Most people with febrile illness due to West Nile virus recover completely, but fatigue and weakness can last for weeks or months. About 1 out of 150 infected people develop a serious, sometimes fatal, illness.

### Lab Exercise -ELISA Simulation

#### Materials

Positive control

Negative control

Patient A –F samples (mimic blood serum)

Simulated antigen, antibody, and chromogen (a colorless chemical compound that can be converted by chemical reaction into a colored compound)

16 microtiter plates

180 plastic pipets

#### Methods

Refer to students' guide provided

#### Expected Results

Record data in the following data table:

Sample	Color	Test Result
Positive control		
Negative control		
Patient A		
Patient B		
Patient C		
Patient D		
Patient E		
Patient F		

\*No Assignment is associated with this lab module. Students' lab notes for results will be evaluated by the instructor.

**Lab 12 Discussion Sessions of Research Papers**  
**Research Project (Continued)**  
Na Xu

**Objective**

- To be familiar with the discussion sessions of a research paper
- Collect data for the group research project

**Introduction**

The discussion session is to interpret and describe the significance of your findings in light of what was already known about the research problem being investigated, and to explain any new understanding or fresh insights about the problem after your research. The discussion will always connect to the introduction by way of the research questions or hypotheses you posed and the literature you reviewed, but it does not simply repeat or rearrange the introduction; the discussion should always explain how your study has moved the reader's understanding of the research problem forward from where you left them at the end of the introduction. This section is often considered the most critical part of a research paper because it most effectively demonstrates your ability as a researcher to think critically about an issue, develop creative solutions to problems based on the findings, and formulate a deeper, more profound understanding of the research problem you are studying. The discussion section is where you **explore the underlying meaning of your research**, its possible implications in other areas of study, and the possible improvements that can be made in order to develop the concerns of your research further.

This is the section where you need to present **the importance of your study** and how it may be able to contribute to and/or fill existing gaps in the field. If appropriate, the discussion section is also where you state how the findings from your study **revealed new gaps** in the literature that had not been previously exposed or adequately described. A good discussion section includes an **analysis of any unexpected findings**. If more than one unexpected finding emerged during the study, describe each in the order they appeared as you gathered the data.

The content of the discussion section of your paper most often includes

1. **Explanation of results:** briefly reiterate for your readers the research problem or problems you are investigating and the methods you used to investigate them, then describe the major findings of the study. You should write a direct, declarative, and succinct proclamation of the study results. You can comment on whether or not the results were expected and present explanations for the results; go into greater depth when explaining findings that were unexpected or especially profound. If appropriate, note any unusual or unanticipated patterns or trends that emerged from your results and explain their meaning. When writing the discussion section, you should carefully consider all possible explanations for the study results, rather than just those that fit your prior assumptions or biases.



2. **References to previous research:** compare your results with the findings from other studies, or use the studies to support a claim. This can include re-visiting key sources already cited in your literature review section, or, saving them to cite later in the discussion section if they are more important to compare with your results than being part of the general research you cited to provide context and background information.
3. **Identification of potential limitations and weaknesses:** comment on the limitations and weaknesses of the research and their relative importance in relation to your overall interpretation of the results and, if necessary, note how the limitations and weaknesses may affect the validity of the findings. Avoid using an apologetic tone; however, be honest and self-critical. Describe the generalizability of your results to other situations, if applicable to the method chosen, then describe in detail problems you encountered in the method(s) you used to gather information. You should also consider any unanswered questions or issues your study did not address.
4. **Deduction:** based on your research, you can propose a more general claim or possible conclusion arising from the results. You can also explain how the results can be applied more generally. For example, describing lessons learned, proposing recommendations that can help improve a situation, or recommending best practices. You should explain the meaning of the findings and why they are essential.
5. **Making suggestions for future studies:** although your study may offer important insights about the research problem, other questions related to the problem likely remain unanswered or arise because of your study. You should make suggestions for further research in the discussion section.

### **Exercise - Research Project Continued**

Students will work in groups to continue their research in investigating the impact of increased dietary sugar on *Drosophila*'s weight, life span, and survivability of F1 progeny. After two weeks of setting up the crosses, data should be collected and analyzed.

#### Materials

Cross set up in lab 10

Stereo microscopes

Forceps

Fly nap

#### Experimental Procedures:

1. Students will use fly nap to put F1 progeny sleep and analyze the following aspects of F1 progeny:
2. Count the number of flies in the F1 progeny.
3. Weigh the total mass of the entire F1 progeny. Calculate the average mass of an F1 progeny fly by the following equation:  
$$\text{Average mass (g)} = \text{total mass (g)} / \text{number of flies}$$

4. The survivability of F1 progeny flies can be evaluated by counting the total number of F1 progeny.
5. Comparing the average mass of F1 progeny grew up in the food vials containing 0, 1%, 4%, 10%, or 15% sucrose and record your result in Table 1:

**Table 1 Average Mass and Survivability of F1 Progeny Flies**

	Flies grown with normal food	Flies grown with 1% additional sucrose	Flies grown with 4% additional sucrose	Flies grown with 10% additional sucrose	Flies grown with 15% additional sucrose
Total Mass (g)					
Number of flies					
Average mass (g)					

### **Assignments-Data Analysis**

Students will draft the **Results, Conclusion and Discussion sessions** of their research project. The minimum word requirement for this session is 600 words.

## References

- Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. Chapter 24, The Adaptive Immune System. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK21070/>
- Celniker SE, Rubin GM. The *Drosophila melanogaster* genome. Annu Rev Genomics Hum Genet. 2003;4:89-117. doi: 10.1146/annurev.genom.4.070802.110323. PMID: 14527298.
- Gupta RM, Musunuru K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. J Clin Invest. 2014 Oct;124(10):4154-61. doi: 10.1172/JCI72992. Epub 2014 Oct 1. PMID: 25271723; PMCID: PMC4191047.
- Hales K, KoreyC, Larracuenta A, Roberts D, Genetics on the Fly: A Primer on the *Drosophila* Model System, *Genetics*, Volume 201, Issue 3, 1 November 2015, Pages 815–842, <https://doi.org/10.1534/genetics.115.183392>
- Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014 Jun 5;157(6):1262-1278. doi: 10.1016/j.cell.2014.05.010. PMID: 24906146; PMCID: PMC4343198.
- Tebas P, Stein D, Tang WW, Frank I, Wang SQ, Lee G, Spratt SK, Surosky RT, Giedlin MA, Nichol G, Holmes MC, Gregory PD, Ando DG, Kalos M, Collman RG, Binder-Scholl G, Plesa G, Hwang WT, Levine BL, June CH. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. N Engl J Med. 2014 Mar 6;370(10):901-10. DOI: 10.1056/NEJMoa1300662. PMID: 24597865; PMCID: PMC4084652.
- Steneck N and Zinn D. ORI introduction to the responsible conduct of research. 2006. <https://ori.hhs.gov/education/products/RCRintro/index.html>
- Susan E. Celniker and Gerald M. Rubin. The *Drosophila Melanogaster* Genome. Annual Review of Genomics and Human Genetics 2003 4:1, 89-11
- Gene structure* - Wikipedia. (2022, June 2). Gene Structure - Wikipedia; en.wikipedia.org. [https://en.wikipedia.org/wiki/Gene\\_structure](https://en.wikipedia.org/wiki/Gene_structure)
- Alternative splicing* - Wikipedia. (2022, June 23). Alternative Splicing - Wikipedia; en.wikipedia.org. [https://en.wikipedia.org/wiki/Alternative\\_splicing](https://en.wikipedia.org/wiki/Alternative_splicing)

